COMPUTING edge

- Software Engineering
- Blockchain Technology
- User Experience (UX)
- Education

JANUARY 2026

www.computer.org

IEEE COMPUTER SOCIETY | 80th ANNIVERSARY CELEBRATION | IEEE

COMPUTING edge

IEEE COMPUTER SOCIETY computer.org

Printed with inks containing soy and/or vegetable oils

Circulation: *ComputingEdge* (ISSN 2469-7087) is published monthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 2001 L Street NW, Suite 700, Washington, DC 20036.

Postmaster: Send address changes to *ComputingEdge*-IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Printed in USA.

JANUARY 2026 • VOLUME 12 • NUMBER 1

COMPUTING edge

Software Engineering

Blockchain Technology

User Experience (UX)

Education

Departments

Magazine Roundup

The IEEE Computer Society's lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

Computer

When the Code Autopilot Breaks: Why Large Language Models Falter in Embedded Machine Learning

This article, featured in the November 2025 issue of *Computer*, presents an empirical investigation of failure modes in large language model (LLM)-powered embedded machine learning pipeline, based on an autopilot framework that orchestrates data preprocessing, model conversion, and on-device inference code generation. Though grounded in specific devices, the authors' study reveals broader challenges in LLM-based code generation.

Computing in Science & Engineering

Combining Automatic Prediction Strategies Using Out-of-Sample Evaluations

Sometimes, it is necessary to predict hundreds or thousands of time series quickly and efficiently. Currently, there are computational automations (automatic prediction strategies) of some forecasting methodologies that can perform this task relatively easily. However, it is not possible to know in advance which of these automations should be employed, and once one has been chosen, all the series in the set must be predicted using the same forecasting methodology. The authors of this July–September 2025 *Computing in Science & Engineering* article discuss the aspects that should be considered to propose a combination of previous existing automations.

IEEE Annals of the History of Computing

How CAD Became Universal

From its origins in the 1960s, computer-aided design (CAD) was initially developed by airplane and automobile manufacturers to solve tough problems. In 1969, entrepreneurs raised venture capital to begin development of commercial CAD systems to be sold to firms that couldn't afford to write their own software. These systems cost about $150,000 per workstation. Most companies in manufacturing and construction industries couldn't afford these prices. This article, featured in the July–September 2025 issue of *IEEE Annals of the History of Computing*, traces the history of how CAD systems evolved from their high-priced origins to become universal across all industries that employ design engineers.

IEEE Computer Graphics and Applications

The State of Single-Cell Atlas Data Visualization in the Biological Literature

Recent advancements have enabled tissue samples to be profiled at the unprecedented level of detail of a single cell. While the problem of cellular data visualization is not new, the size, resolution, and heterogeneity of single-cell atlas datasets present challenges and opportunities. The authors of this September/October 2025 *IEEE Computer Graphics and Applications* article survey the usage of visualization to interpret single-cell atlas datasets by assessing over 1800 figure panels from 45 biological publications. This report intends to be a foundational resource for the visualization community as atlas-scale single-cell datasets are emerging rapidly with aims of advancing understanding of biological function in health and disease.

IEEE Intelligent Systems

Semantic Map Construction Under Complex Weather Scenarios

Semantic maps provide information on road elements, which is crucial for ensuring driving safety. However, previous methods mostly focus on normal weather conditions, neglecting the challenges of scene feature extraction caused by image degradation in complex weather scenarios. To address these challenges, the authors of this article featured in the September/October 2025 issue of *IEEE Intelligent Systems* propose a normal weather scene-guided complex weather scene map construction network (NCMC-Net).

IEEE Internet Computing

Characterization of Probabilistic Structure of Internet Traffic During COVID-19: A Study Based on MAWI Data

The COVID-19 pandemic greatly affected all aspects of human life, including operations of offices, businesses, industries, and educational institutions. With an increased shift to online work, changes in Internet traffic characteristics are inevitable. This article from the May/June 2025 issue of *IEEE Internet Computing* presents a study based on Measurement and Analysis on the WIDE Internet (MAWI) data characterizing traffic in terms of multimodal and unimodal probability distributions. These findings are likely to be useful to Internet service providers in managing future traffic.

IEEE micro

Enabling Sustainable Cloud Computing With Low-Carbon Server Design

To combat climate change, carbon emissions from hyperscale cloud computing can be reduced. Compute servers cause most of the general-purpose cloud's emissions. Thus, the authors of this article, featured in the July/August 2025 issue of *IEEE Micro*, were motivated to design carbon-efficient compute server stock keeping units (SKUs), or GreenSKUs, using recently available low-carbon components. This work is the first to demonstrate and quantify how carbon-efficient server designs translate to measurable cloud-scale emissions reductions, enabling meaningful contributions to cloud sustainability goals.

IEEE MultiMedia

Robust and Multilayer PowerPoint Watermarking for Source Tracing

Hybrid work with online circulation of office documents becomes the new norm while also giving rise to security risks like information leakage. Digital watermarking technology is an effective method for tracking multimedia data dissemination and ensuring authentication. The authors of this July–September 2025 *IEEE MultiMedia* article propose a robust and blind multilayer PowerPoint document watermarking scheme based on slide format and zero-width characters.

IEEE Pervasive Computing

Mobile Systems | Ubiquitous Computing | Internet of Things

Multifunctional Electrodes for Signal Soil Measurements: Benchmark With ML-Based Algorithms

Soil monitoring is a crucial issue for sustainable field and agricultural management. This article, featured in the July–September 2025 *IEEE Pervasive Computing* issue, explores the performance of machine learning models in classifying soil types under varying moisture levels using wire–plate and plate–plate sensor configurations. The study highlights the importance of sensor design, model selection, and environmental factors in optimizing soil classification accuracy.

IEEE SECURITY & PRIVACY

Employee Privacy Protection in Enterprise Generative Artificial Intelligence Deployment

In this article, featured in the September/October 2025 issue of *IEEE Security & Privacy*, the author proposes a multilayer privacy protection framework to mitigate risks from enterprise generative artificial intelligence (AI) adoption. By integrating prevention, monitoring, response, and governance mechanisms, the framework ensures ethical AI deployment while safeguarding employee privacy rights and organizational productivity.

IEEE Software

Tool: QUIET: A Tool for Sampling-Based Quantum Noise Error Mitigation

Quantum noise poses a significant obstacle in fully realizing the potential of quantum computing. Although existing quantum error mitigation techniques focus mainly on correcting errors in expectation value-based outputs produced by quantum software, there are fewer solutions for mitigating errors in sampling-based outputs, which are essential for algorithms like Shor's and Grover's. In this article, from the November/December 2025 issue of *IEEE Software*, the authors present QUIET, a tool that implements the state-of-the-art sampling-based error mitigation technique QLEAR to reduce noise in the sampling-based output of quantum software.

IT Professional

Artificial Intelligence in the Middle East and Africa: Needs and Requirements

In this article, featured in the September/October 2025 issue of *IT Professional*, the authors discuss the unique set of needs and opportunities in the Middle East and Africa (MEA) region for generative artificial intelligence (AI), driven by economic diversification, digital transformation, and social challenges. Key sectors such as health care, education, finance, and governance require AI-driven solutions tailored to linguistic, cultural, and infrastructural nuances. A region-specific approach to generative AI can enhance economic growth, societal well-being, and sustainable development in the MEA.

Software Engineering: Fundamentals and Future

The field of software engineering is constantly evolving. To keep up, engineers must maintain a strong grasp of the core principles of software engineering while staying up-to-date with new directions in the field. This issue of *ComputingEdge* investigates top software principles over time, the use of large language models (LLMs) in the bug repair process, and the integral role of collaboration in software engineering. The articles also explore changes and standardization practices in blockchain technology and user experience (UX) of artificial intelligence (AI) based on different LLMs. The issue concludes with a discussion of the perception and uses of AI in education.

It is important for engineers to understand the principles that have shaped software engineering since its inception as well as topics and principles that pertain to future directions in the field. *IEEE Software* article "Software Principles" takes a deep dive into the essence of software technology over time by outlining top software principles. The authors of "Can AI Fix Buggy Code? Exploring the Use of Large Language Models in Automated Program Repair" from *Computer*, present a study that reviews the current relationship between human engineers and LLMs in the bug repair process as well as the potential for LLMs to fix bugs without human intervention. The *IEEE Software* article "Software Development Is a Team Sport," argues that teamwork is essential to the engineering process.

Blockchain technology and standardization have changed significantly in recent years. In "From Crisis to Comeback: The Evolution and Resurgence of Blockchain Post-2022 Crypto Winter" from *Computer*, the author explores how blockchain technology is recovering and adapting following the 2022 crypto market downturn. *Computer* article "Blockchain Standardization in Practice: Contrasting European Union and U.S. Approaches" categorizes blockchain standards and contrasts the European Union's and the United States' regulatory approaches in standardization.

Users have varying experiences with artificial intelligence based on which LLM they use and how they use it. In "Gemini Versus ChatGPT and DeepSeek: Much Ado About Crawling" from *Computer*, the author compares ChatGPT-4o and DeepSeek-R1, revealing key differences in speed, consistency, and UX. The authors of "Necessary but Not Perfect: Changes in AI Perception at a Large University," from *IT Professional*, show how the perception of AI has changed on one university campus through AI literacy events, conferences, and presentations. Survey results suggested an increase in AI understanding and usability, though ethical concerns remained the same.

Communication is an essential part of education. It plays an important role in how people learn as well as how AI can be effectively designed and implemented to aid education. In "Foundation Models for Education: Promises and Prospects" from *IEEE Intelligent Systems*, the authors explain how foundation models can promote education, as well as the risks and opportunities of AI overreliance and creativity.

EDITOR: Christof Ebert, Vector Consulting Services, christof.ebert@vector.com

This article originally appeared in IEEE Software vol. 41, no. 2, 2024

DEPARTMENT: SOFTWARE TECHNOLOGY

Software Principles

Christof Ebert and Alan M. Davis

FROM THE EDITOR

This is the 100th installation of the technology column of *IEEE Software* magazine. Rather than providing statistics, we will investigate the essence of software technology over all these years—and into the future. Which are the relevant software principles? What principles matter for software practitioners? The article provides the top-10 principles of all times based on a survey. Another top-10 list with novel principles covers more recent evolution in software engineering. I look forward to hearing from you about this column and the technologies that matter most for your work.—*Christof Ebert*

Somewhere at this moment, a software *project* is failing. Somewhere at this moment, a software *product* is failing. While a project failure impacts a company, a product failure also impacts users and maybe even societies. One reason projects and products fail is that underlying software principles are not being observed.[1,2] Projects are often started without knowing the major requirements and having no means to balance business needs versus technical demands. Many products do not provide the necessary quality, such as cybersecurity, resilience, reliability and functional safety. For instance, artificial intelligence (AI) is often deployed without understanding its implications and not being able to sufficiently test. While there are no silver bullets for software development, knowing about software principles and practicing (!) them will help.

> *PROJECTS ARE OFTEN STARTED WITHOUT KNOWING THE MAJOR REQUIREMENTS AND HAVING NO MEANS TO BALANCE BUSINESS NEEDS VERSUS TECHNICAL DEMANDS.*

SOFTWARE PRINCIPLES

Software is not tangible, and so physical laws do not form a suitable foundation. To avoid biased judgment, software engineering has had to evolve its principles based on empirical studies, practitioner guidance, and case studies.[3,4] Many software accidents have their root cause in not following such basic software principles.

- A single-point failure in the maneuvering characteristics augmentation system (MCAS) in the Boeing 737 MAX caused two air accidents killing 346 people, demonstrating failures to thoroughly specify and design software, and to train its users.
- The repeated ransomware attacks on software systems worldwide demonstrate vulnerabilities in software.
- An unpredicted and uncaught overflow error caused human flight controllers to have to destroy the Ariane 5 missile immediately after launch, showing that reuse means more than simple copy-paste.
- The Mars Climate Orbiter crashed into Mars due to a miscommunication between two

Digital Object Identifier 10.1109/MS.2023.3343184
Date of current version: 22 February 2024

programmers concerning units of measurement for a variable: one thought pounds, the other thought international measurement system.

When a bridge or building collapses, investigators try to determine what went wrong. Usually, it is because a builder failed to comply with a building code (i.e., a set of rules, or principles, to follow during construction), or an inspector failed to locate a physical deterioration. When software fails, it is usually because a software engineering organization failed to comply with a principle.

What are principles? What makes a principle a principle? Principles help us to stay on course in what we are doing. For instance, agile is based on principles that facilitate flexibility, such as value and ownership. Yet, agile is also an example, where a hype eventually drives misunderstanding. Today, agile is often misinterpreted as "everything is possible" or "little process." This leads to errors and rework, exacerbated by distributed teams. Systematic working with principles does not mean formalism or even dogmatism.[3,5]

Principles provide guidance in evolving and unknown territory. They are generally valid and tend to be relatively abstract. They keep us on track in changing environments and with challenging constraints. Our way of working can be pragmatically adapted by adhering to unchanging principles. Such need for flexibility was one of the major drivers for agile behaviors. Not everything can be planned to the detail. We must plan for uncertainties and risks, which means having not only one plan, but a set of variables which we can adjust if necessary—while still achieving the major business needs.

Principles are derived from and confirmed by experience and knowledge. They do not depend on a specific technology, notation, or paradigm. A paradigm can support given principles such as object-orientation enhanced information hiding. Some software principles were coined in the early years of software engineering, over 50 years ago—and are still valid and used. Figure 1 shows some software principles in context. Examples include the following:

- abstraction
- structuring
- binding and coupling
- hierarchization
- modularization
- data hiding
- localization
- self-explanation.

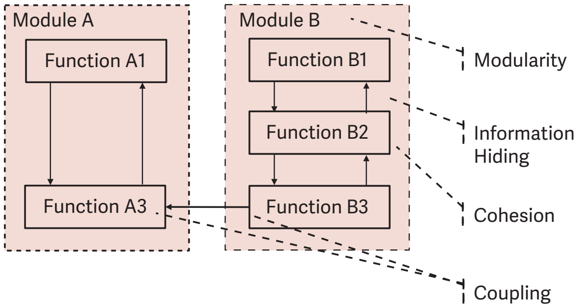


FIGURE 1. The principles of modularity, information hiding, cohesion, and coupling.

SOME SOFTWARE PRINCIPLES WERE COINED IN THE EARLY YEARS OF SOFTWARE ENGINEERING, OVER 50 YEARS AGO—AND ARE STILL VALID AND USED.

Such principles played an important role in software development, especially for specification, design, and implementation. "Goto considered harmful" by Dijkstra[6] helps us to preserve locality and structuring. "Design for change" by Parnas[7] facilitates long life for software. These early software principles are still extremely valuable and valid today.

Engineering disciplines evolve based on proven principles. Electrical engineering (EE) shows us this evolution path.[5] For centuries there was no discipline called EE and even associated crafts were rather magic. Reducing field observations to what was essence, as opposed to accidents, established its principles. A completely new scientific branch of physics was shaped by scientists such as Kirchhoff, Faraday, and Maxwell. Derived EE principles transformed the underlying physical laws to an engineering discipline. Ever since, new theoretical insights such as quantum hall effect that swapped over from physics were deduced to principles and further on used by engineers to provide innovative artifacts for the use of humanity.

A good example of such chains from theory to practice is the generalization of Kirchhoff's laws to the theory of circuits and then to CAD tools that allow for easy design, test, and simulation of new circuits. Resulting problems from daily applications of circuit theory are continuously played back and forth to the theoretical branch of EE to obtain applicable solutions. The tools in turn adapt and provide these theoretical results almost immediately in a way that is useful for practitioners.

TO EVALUATE THE INITIAL 201 PRINCIPLES AND THEIR IMPACT ON TODAY'S SOFTWARE ENGINEERING DISCIPLINE, WE ASKED LEADING PRACTITIONERS AND RESEARCHERS ON THEIR PERCEPTIONS.

Often, practitioners and students ask for what really matters in software engineering. They want to know about key principles, similar to what we have in other engineering disciplines. One early attempt to summarize software engineering principles was the book *201 Principles of Software Development* by Alan Davis.[3] When compiling the principles, he covered the entire lifecycle, not just design as was the case with earlier attempts. He even introduced a section on software evolution, which at that time was hardly thought about. Twenty-five percent of his principles were on design and code, 25% on management, 20% on requirements and test, 15% on software control and evolution, and a similar amount of a more generic character. With the knowledge and experience of the past 30 years, let us look at what is still valid and what might be new.

201 SOFTWARE PRINCIPLES REVISITED

To evaluate the initial 201 principles and their impact on today's software engineering discipline, we asked leading practitioners and researchers on their perceptions. The experiment used a confidential survey from experts in the Americas, Asia, and Europe. Their common background was software engineering, spanning teaching, research, practice, and managing companies.

The result was a ranked list of those existing 201 principles plus a list of new recommended principles. Some of the existing principles were no longer considered applicable. Yet many prevailed, and surprisingly few new rules appeared. Table 1 shows the 10 existing principles that received the most votes. The first column is the original numbering of the principles in Davis,[3] the next column the wording in Davis,[3] and the third column provides some discussion that was triggered by the survey.

Keeping these timeless principles in mind, let us look, what is missing—from today's perspective. We asked the same audience in the survey to specify what new principles they consider most relevant in their environment. Table 2 lists the novel top-ten principles. Not all are entirely new. For instance, what today is called "technical debt" was the entropy principle in the initial list of 201 principles. Marketing lives on hypes. That also applies occasionally for software engineering, as we know from the technology hype-cycle.

Trust in software products is decreasing. One reason is that complexity is growing faster than competencies. But this is a weak excuse. We should rather strive for a better understanding of what we are doing as software practitioners and how to ensure that we deliver the right product with the right quality. In the fast-evolving landscape of software development, adherence to principles serves as a guiding force that shapes robust, maintainable, and scalable solutions. Software practitioners often navigate a complex set of technologies, frameworks, and methodologies. Amidst this complexity, adherence to principles will guide the decision-making process and thus influence the quality of software systems.

Principles must be made pragmatic to be used (see "Putting Traceability from Principle to Practice" for a case study). This means to connect principles with methods that practitioners can and will apply. Figure 2 shows how "principles" relate to other fundamental terms that constitute software engineering as a science.

Not practicing principles creates major risks along the lifecycle. Take cybersecurity. Increasingly generative AI (GenAI) tools and platforms are used to facilitate reuse and accelerate deliveries. Yet GenAI

TABLE 1. Timeless top-10 of the original 201 principles.

Number	Principle	Relevance and discussion
14	Build in small increments.	The high runner across all groups is this very basic principle to control complexity, namely divide and conquer. This agile principle holds until today and applies to practically all software. The agile manifesto has collected towards the late nineties such principles to ease projects.
37	Take responsibility.	Software practitioners and managers must take responsibility. They need to understand a problem, its risks, and make decisions for which they are hold accountable. Also, this is a key agile principle which might be seen as overly general, but especially in software is crucial as there are often many cooks, but nobody to take the helm.
8	Listen to your customer.	This principle addresses the key reason for failure on many projects today.[1,2] Before starting, one must understand the problem and document it. The principle does not say to build what customers say, it does say that you must just listen to them, i.e., don't assume that the customer necessarily needs what he is asking for.
98	Inspect code.	Code inspection might seem awkward and old-fashioned. Yet it ranks high because in the age of many security attacks and a growing amount of AI-generated code, it is relevant that humans stay in control. Inspections should be done with appropriate checklists and with support of tools such as static analysis. Record what you observe and get that back to the developer.
131	People are the key to success.	People matter, a no-brainer for every successful manager. Yet, many managers see people as interchangeable and thus totally replaceable. This is dubbed as managing by Excel. A sure recipe for failure as we see in many postmortems.
181	Keep track of every change.	Changes often happen without being analyzed upfront and recorded. The result is unnecessary complexity, features that nobody uses, and a huge amount of extra cost. Worse yet, the software is of insufficient quality because the changes are rarely regression tested. Document changes and ensure that there is a regression test.
50	Prioritize requirements.	Priorities allow us to partition work. Priorities allow us to focus on where it matters most. It also facilitates triage[8] to resolve the never-ending conflict between marketing or customers (who want it all) and developers (who want schedules and budgets to align with requirements).
7	Give products to customers early.	Similar to "build in small increments," and "listen to your customer," this principle completes the concept by telling us what to do with the knowledge gained. Showing early prototypes helps in overcoming the "I know it when I see it" problem. With complex user interfaces becoming the number one reason of software defects, early feedback improves user experience. Agile methods such as design thinking are built on this principle.
74	Design for change.	Software is subject to change, and developers must prepare for it. Classic software principles such as information hiding and modularity help.[7] Design for change also means documenting requirements and regression tests. Yet today we still see software which is not suited for change. Netscape Navigator is one example where a company with 90% market share lost it all, just because the code was not designed for change.
134	Trust your people.	Knowing that people are the single most relevant success factor, it is important not only to empower and demand ownership, but also to show trust. Give and take. With today distributed teams this principle not only applies to line management but also to team management and distributed working. Trust is built upon commitments that are delivered. Mechanisms such as Scrum and Kanban boards allow to record commitments, so the team knows about who will do what next.

might be misused and trained to insert unwanted code fragments into any code it processes. Such snippets might look innocent but could introduce backdoors, manipulate data, or feed information to external targets. Though this holds for any code reuse, cyber warfare will enter a new stage with AI-based generated code which is hard to understand and test. Using AI and ensuring trust to derived products needs new software principles to verify and validate AI, such as its continuous evolution. Software practitioners need to enhance their competences on the right side of the "V" abstraction to ensure reliability, robustness, and resilience.

"Practicing principles is more important than proving them." The ancient philosopher Epictetus gave us this wisdom. We need principles that we don't constantly question. Asking questions is important, but if you don't have a solid foundation of principles, you will fail because of constant doubts. This is particularly important these days of dramatic changes in society and technology, where we must continuously balance decisions and their impacts. Principles serve as a compass, steering practitioners toward solutions that are not only functional but also sustainable.

SOFTWARE PRACTITIONERS NEED TO ENHANCE THEIR COMPETENCES ON THE RIGHT SIDE OF THE "V" ABSTRACTION TO ENSURE RELIABILITY, ROBUSTNESS, AND RESILIENCE.

With 100 installations of this technology column, we continuously map technology advances with industry experiences and guiding principles. As software practitioners navigate the intricate

TABLE 2. Novel Top-10 Principles that Were Not in the Initial List.

New	New principles	Relevance and discussion
1	Detect defects early to avoid high correction cost.	Late defect detection needs a longer correction cycle than if the defect was detected close to its introduction. This holds specifically for requirements where reviews reduce expensive rework. This principle directly addresses continuous verification.
2	Ask why.	A principle in just two words is rare. This one is well-known from children who never stop asking "why?." Yet it also applies in requirements, such as a new need of a client. Asking why helps to understand reasons.
3	Document with scenarios.	Pictures and stories are much easier to comprehend than long comments and specs. Thinking in scenarios visualizes a usage of a software system. Describing scenarios helps to identify critical exceptions, vulnerabilities, and errors.
4	Balance features, schedules, and budgets.	People who specify or design software can never have enough features. But less is more, as we learned from the Chicago school of architects. We must balance needs with constraints. Doing more needs more time and effort, which many customers and markets are just not willing to pay.
5	Focus on value: reduce accidents, control essence (RACE).	Value is what makes a customer buy our product or service. High value makes the customer say "Wow!." Value is in the eye of the beholder. As developers we must reduce accidents, e.g., features without value, rework, and defects. And we need to control essence, e.g., develop according to priorities and test what is delivered.
6	Control liabilities.	An old principle which was the initial number one in the list of 201 principles was "Quality first." While quality needs balance with other constraints, many products eventually fail due to focus on speed rather than quality. Product liabilities due to project constraints were later called "technical debt."
7	Deploy defensive and robust coding practices	Errors happen. A human typically delivers one error in ten lines of any work products. We typically find half ourselves, but the other half remains. Verification and validation help but are not perfect. To ensure reliable behavior, design and code must be robust and exploit mechanisms to not immediately crash, but at least fail operationally. A watchdog is such a mechanism to avoid silent termination of a program.
8	Continuously grow your competencies.	Software as a discipline is continuously evolving. New technologies, methods, tools, and application domains must be understood and digested. Halftime of knowledge is rarely as short as in software. Software practitioners and (!) managers must continuously learn. Learning is not enough, but also needed is doing. Ask yourself each single evening what you have learned during the day. If it is not clear, read an article like this one and you have increased your competences.
9	Create, maintain, and reassess the business case.	Base your decisions on a tangible business reasoning. Assess and document benefits and cost of requirements, technologies, tools, and so on. Business assumptions change as do user needs, constraints, and markets. Re-evaluate your assumptions every few months. Learn from it and thus improve your judgment.
10	Don't throw good money after sunk cost.	As engineers, we tend to believe that what we have developed might somehow be of value. Wrong, because it is not our judgement but that of markets and customers. If the business outlook of a product or asset becomes meager and unsatisfactory, it is time to compare cost versus benefits to complete versus to terminate. What we have done so far is only of value if there is somebody willing to pay for it.

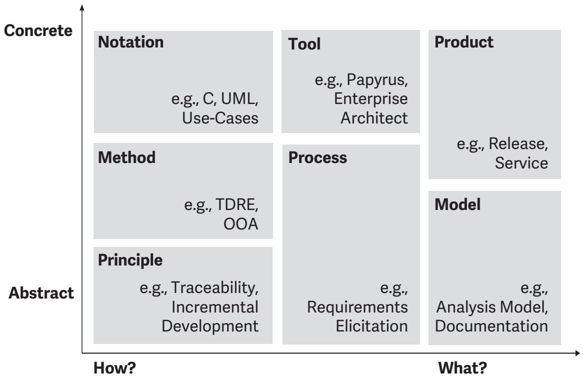


FIGURE 2. Principles versus method, notation, process, tool, product, and process.

technologies, understanding and applying underlying software principles is a key driver for success in building resilient software products with value beyond their initial specification delivery.

REFERENCES

1. "Leading reasons for software project failure." Statista. Accessed: Nov. 24, 2023. [Online]. Available: www.statista.com/statistics/627648/worldwide-software-developer-survey-project-failure/
2. "Why IT projects still fail," CIO Magazine, Oct. 16, 2023. Accessed: Nov. 24, 2023. [Online]. Available: www.cio.com/article/230427/why-it-projects-still-fail.html
3. A. Davis, *201 Principles of Software Development*. New York, NY, USA: McGraw-Hill, 1995.
4. B. Boehm, "Seven basic principles of software engineering," *J. Syst. Softw.*, vol. 3, no. 1, pp. 3–24, Mar. 1983, doi: 10.1016/0164-1212(83)90003-1.
5. C. Ebert, "The road to maturity: Navigating between craft and science," *IEEE Softw.*, vol. 14, no. 6, pp. 77–82, Nov./Dec. 1997, doi: 10.1109/52.636674.
6. E. Dijkstra, "Letters to the editor: Go to statement considered harmful," *Commun. ACM*, vol. 11, no. 3, pp. 147–148, 1968, doi: 10.1145/362929.362947.

PUTTING TRACEABILITY FROM PRINCIPLE TO PRACTICE

Putting principles to practice is not easy. For instance, the traceability principle is well known and taught in every software engineering class. Yet, most software is not connected to requirements and adequate test cases. Requirements are typically developed without even thinking about how they will be tested. Designers will not feel the problem because they are solution minded. Later, a tester looks at the problem description and starts asking about correlations and context. Often this leads toward reopening the requirement with an astounding amount of rework and overheads.

In supporting industry projects, we face the following set of challenges related to traceability:

- Traceability is not maintained, vertical and horizontal.
- Test cases are generated at random, leaving white spots and inefficient repetition.
- Software requirements remain incomplete and incompletely linked to the higher-level system requirements.
- Major quality requirements are not connected to the software design.
- There is unnecessary variance due to copy/paste of code without a clear architecture.

Here are some recommendations for establishing and maintaining traceability.

- *Pragmatism*: Make traceability hands-on, both what to do and how to measure—and then walk the talk. For three decades, traceability has been in all software engineering books and town hall presentations, but not practiced. Often it is perceived by practitioners as "management fad" because there is no hands-on guidance. We train it to architects, and they tell us that they see the value, but nobody really cares.
- *Key rules*: Provide guidance for developers as simple key rules. It is of no help to demand traceability if there are no simple rules what to do, and how to decide in conflicts of insufficient time budget or conflicting design decisions. Examples of rules are as follows: Include one test case per requirement; no "nomadic" components which are not linked to requirements; prioritize relevant quality requirements and allocate them to software. Whatever the rules, ensure it can be practiced throughout the lifecycle.
- *Connect principle with methods*: Establish a simplified and consistent guidance and integration of individual processes. The method of test-driven requirements engineering (TDRE) allows us to systematically develop test cases together with the respective requirements and thus achieve basic traceability. As a minimum, the requirement should be traced to the sunny day scenario requirement. It might be enhanced with test cases for critical correlations of requirements and negative requirements such as misuse cases in cybersecurity. Practicing the principle of traceability facilitates consistency, quality, and testability. It yields a minimum viable set of test cases which serve for regression test.
- *Key Performance Indicators*: Establish measurements for impacts of insufficient traceability and visualize in product reviews. The major challenge of most projects is that there are too many things to do, and no transparent decision making based on measurements. Traceability will only be done if there is a budget which means that the risks of insufficient traceability and mitigation is monetarized.

Traceability as a principle is not self-sustained but a culture change which takes time and budget—based on explicitly monetarizing the impact and benefits.

7. D. Parnas, "Designing software for ease of extension and contraction" in *Proc. 3rd Int. Conf. Softw. Eng.*, May 1978, pp. 264–277, doi: 10.5555/800099.803218.
8. A. Davis, "The art of requirements triage" *Computer*, vol 36, no. 3, pp. 42–49, Mar. 2003, doi: 10.1109/MC.2003.1185216.
9. F. Calefato and C. Ebert, "Agile collaboration for distributed teams [Software Technology]," *IEEE Softw.*, vol. 36, no. 1, pp. 72–78, Jan./Feb. 2019, doi: 10.1109/MS.2018.2874668.

CHRISTOF EBERT is the managing director of Vector Consulting Services, 70499 Stuttgart, Germany. Contact him at christof.ebert@vector.com.

ALAN M. DAVIS is president and CEO of Offtoa, Inc., Westminster, CO 80021 USA. He is a previous editor-in-chief of *IEEE Software*. Contact him at aldavis171@gmail.com.

EDITOR: Hsiao-Ying Lin, IEEE Member, hsiaoying.lin@gmail.com

DEPARTMENT: ARTIFICIAL INTELLIGENCE/ MACHINE LEARNING

This article originally appeared in Computer vol. 58, no. 7, 2025

Can AI Fix Buggy Code? Exploring the Use of Large Language Models in Automated Program Repair

Lan Zhang, Northern Arizona University

Anoop Singhal, National Institute of Standards and Technology

Qingtian Zou, University of Texas Southwestern Medical Center

Xiaoyan Sun, Worcester Polytechnic Institute

Peng Liu, The Pennsylvania State University

This article reviews the current human–large language models collaboration approach to bug fixing and points out the research directions toward (the development of) autonomous program repair artificial intelligence agents.

The field of software engineering has witnessed a paradigm shift with the advent of large language models (LLMs). These sophisticated artificial intelligence (AI) systems have demonstrated remarkable versatility across various software development tasks, including code generation, bug detection, and code review.[1,2,3] The potential of LLMs to revolutionize software development practices has sparked broad interest within both academic and industry circles, prompting a surge of research into their capabilities and limitations.

A recent breakthrough in this domain came with the introduction of Devin, an LLM-powered AI system capable of autonomously completing 13.8% of real-world coding tasks.[4] These tasks encompass a range of complex operations, from diagnosing and fixing bugs to conducting comprehensive code reviews. However, the relatively modest success rate of 13.8% in real-world scenarios raises a critical question that forms the core of our investigation: Are we truly prepared to leverage LLMs for repairing buggy complex programs? This question is not merely academic but has far-reaching implications for the future of software development and maintenance practices.

To address this fundamental quest, our study focuses on two modes of LLM-supported program repair:

- *Human–LLM collaboration:* This approach examines the synergistic relationship between human software engineers and LLMs in the bug repair process.[5] It encompasses both interactive, dialogue-based methodologies and more integrated solutions, such as real-time code completion and suggestion systems.
- *Autonomous AI agent repair:* This mode investigates the potential for LLMs to independently identify and rectify bugs without direct human intervention, representing a more ambitious vision of automated program repair.

By examining the efficacy of LLMs across diverse programming contexts—for example, C/C++, Java, and Python—we aim to provide a nuanced understanding of their current capabilities and limitations in addressing complex software bugs. Our findings reveal a nuanced landscape of LLM-supported program repair. For the human–LLM collaboration

Digital Object Identifier 10.1109/MC.2025.3527407
Date of current version: 27 June 2025

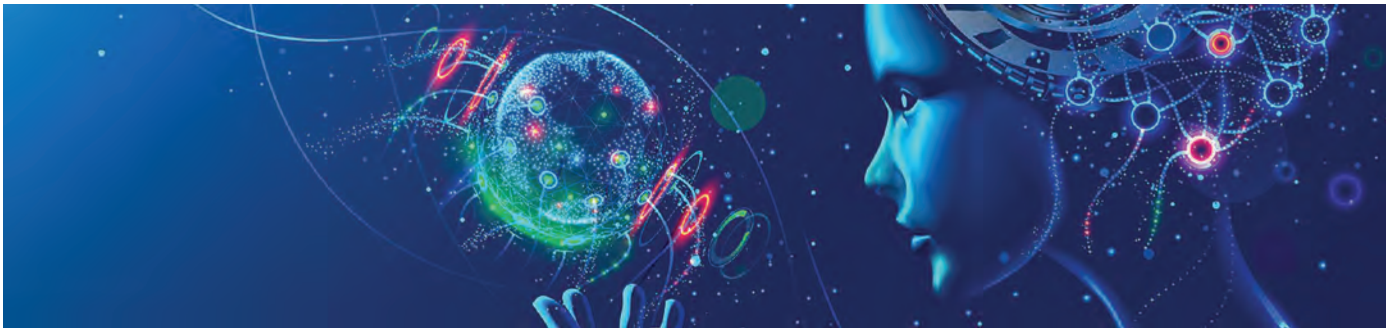


TABLE 1. LLM-based program repair across languages and methodologies.

	Dataset type	Methodologies	Article
C/C++	Synthetic programs	Human-supported dialogue	Yang et al.,[7] Pearce et al.[8]
	Real-world projects	Human-supported dialogue	Zhang et al.,[2] Bajpai et al.,[6] Pearce et al. ,[8] Kulsum et al.[9]
Java	Synthetic programs	Human and static tools supported dialogue	Kang et al.,[10] Xia and Zhang,[11] Wadhwa et al.[12]
	Real-world projects	Human-supported dialogue	Kulsum et al.[9]
Python	Synthetic programs	Human and static tools supported dialogue	Wadhwa et al.[12] Lemieux et al.,[13] Cao et al.[14]
	Real-world projects	Human and static tools supported dialogue	Parasaram et al.,[15] Jimenez et al.[16]

mode, we observed that results could be significantly improved when humans provide additional contextual knowledge. This includes information about variable contexts, relevant data structures, related functions, and even the underlying logic of the code. This synergy between human expertise and LLM capabilities shows promise for enhancing bug repair processes in complex software systems. In contrast, the autonomous AI agent repair mode presents a more challenging frontier. Our research indicates that we are still far from achieving reliable automatic code repair using LLMs alone. The complexity of real-world software systems, coupled with the nuanced understanding required for effective bug repair, continues to pose significant challenges for fully autonomous LLM-based solutions.

HUMAN–LLM COLLABORATION

GitHub Copilot's ROBIN system represents a significant advancement in human–LLM collaboration for debugging.[6] It uses multiple AI agents to analyze code context, exception information, and user queries, guiding developers through systematic debugging steps. ROBIN leverages LLMs as reasoning engines to provide interactive and collaborative debugging assistance through a chat-based interface. It analyzes exception information, code context, and user queries, guiding developers through a series of steps to explore potential hypotheses, gather more information, and utilize IDE debugging tools to fix issues. This industrial work demonstrates the potential for more effective collaboration between developers and AI in software debugging tasks.

To understand the current state and potential of human–LLM collaboration in program repair, we conducted a comprehensive review of existing research across multiple programming languages and methodologies. Table 1 summarizes our findings, categorizing studies based on programming language (C/C++, Java, Python), dataset type (synthetic programs and real-world projects), and methodology.

Our analysis reveals a clear trend across all three programming languages: The performance of

DISCLAIMER

Commercial products are identified to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the identified products are necessarily the best available for the purpose.

EXAMPLE 1: SYNTAX DIFFERENCE IN JAVA AND C++.

Java:

```
String str = "Hello";
List<Integer> numbers = new
ArrayList<>();
System.out.println(str.length());
```

C++:

```
std::string str = "Hello";
int* numbers = new int(5);
std::cout « (*numbers) « std::endl;
delete numbers;
```

LLM-assisted repair techniques tends to decrease as the complexity of the dataset increases.[2,7,8,9,15] For example, Yang et al.[7] and Pearce et al.[8] investigated human-supported dialogue approaches with synthetic programs, achieving remarkably high success rates: up to 100% in some cases. However, when this methodology was extended to real-world projects by Zhang et al.,[2] Kulsum et al.,[9] and again by Pearce et al.[8] the performance dropped dramatically to less than 20%.

While the general trends are consistent across C/C++, Java, and Python, some language-specific nuances emerged. For instance, as depicted in "Example 1: Syntax Difference in Java and C++," Java manages references to objects without explicit pointers, while C++ allows direct memory manipulation through pointers ("str"). Moreover, Java employs automatic memory management through garbage collection, where "numbers" is automatically deallocated when it's no longer referenced or goes out of scope. In C++, we must manually allocate memory with "new" and then explicitly deallocate it with "delete" to prevent memory leaks. Similar to Java, Python uses automatic memory management. Its dynamic typing and high-level abstractions can simplify certain programming tasks, potentially making some types of repairs more straightforward. For example, Python shows the highest success rate at 38.80%,[16] while C/C++ lags behind at 16.5%.[2] The lower performance in C/C++ can be partially attributed to the complexity introduced by manual memory management and pointer manipulation.

Our analysis of the human–LLM collaboration in program repair leads to one key conclusion: Human expertise continues to play a critical role in the bug repair process. Results improve substantially when humans provide additional contextual knowledge.[2,5]

- *Context of variables:* Understanding the context of variables is crucial for LLMs in program repair for several reasons. The scope of a variable, whether it's global, local, or class-level, determines where it can be accessed and modified. LLMs need to understand this to avoid introducing bugs by incorrectly accessing or modifying variables. Knowing the range of possible values a variable can take helps in identifying potential edge cases or unexpected inputs that could lead to vulnerabilities. Understanding how a variable is typically used within the code, such as a loop counter, a flag, or to store intermediate results, helps LLMs generate more appropriate and context-aware fixes. Tracking how the value of a variable changes throughout the program's execution is essential for identifying the root cause of bugs and proposing effective solutions. In dynamically typed languages, inferring the type of a variable from its usage context is crucial for generating type-safe patches.
- *External elements:* Knowledge of external functions, data structures, and variables is vital for LLMs in program repair. LLMs need to understand the correct usage of external application programming interfaces, including function signatures, return values, and potential side effects. For languages with manual memory management, understanding how external functions allocate and deallocate memory is crucial for preventing memory leaks and buffer overflows. Knowledge of how external functions report errors, such as through return codes or exceptions, is necessary for implementing proper error checking and handling in patches. Understanding whether external functions

are thread-safe is important when generating patches for multithreaded applications.

- *Logic of the vulnerable code:* Comprehending the logic of vulnerable code is essential for effective program repair. Understanding what the code is supposed to do is crucial for ensuring that patches fix the vulnerability without breaking the intended functionality. Analyzing the control flow of the vulnerable code helps in identifying potential logical errors or improper handling of edge cases. Understanding how different parts of the code depend on each other's outputs helps in ensuring that patches maintain the correct data flow. Grasping the underlying algorithms used in the code is necessary for proposing optimizations or alternative implementations that might resolve the vulnerability. Identifying implicit security assumptions in the code, such as trust boundaries and input validation, is important for generating patches that enhance the overall security posture.

In "Case Study 1: Understanding the Range of a Variable," providing comprehensive information about the vulnerable code's logic and the relationship between the variables *p*-> tokenpos and *p*-> tokenlen greatly helps LLMs generate effective patches. This additional knowledge should involve explaining the fundamental logic behind the vulnerability, emphasizing the potential disparity between *p*-> tokenpos and *p*-> tokenlen, and clarifying the proper boundary conditions. Specifically, elucidating that *p*-> tokenbuf is capable of holding *p*-> tokenlen elements, making the maximum index *p*-> tokenpos should be *p*-> tokenlen−1.

AI AGENT AUTOMATIC REPAIR

Recent advancements in AI have led to the development of increasingly sophisticated coding assistants, such as Codeium,[17] Devin,[4] Cursor,[18] Magic,[19] Replit,[20] and Cody.[21] Devin AI represents a significant advancement in AI-powered software engineering, demonstrating impressive capabilities in autonomously resolving GitHub issues. This system can build and deploy applications end-to-end, encompassing tasks, such as project cloning, exploration of structures related to vulnerable functions, test case updating based on compiler error messages, generation of new test cases through brute force methods, and bug identification and repair. Devin's ability to resolve 13.8% of issues in the SWE-bench benchmark,[22] outperforming GPT-4 by a factor of three, is a notable technical achievement. While tools like GitHub Copilot, Codeium, and Cody primarily focus on code completion and generation, their underlying technologies contribute to the broader field of automatic code repair. These systems leverage LLMs trained on vast corpora of code, enabling them to understand code context and suggest fixes for common errors. However, it is crucial to contextualize this success within the broader landscape of software development. While capabilities of automatic AI agents are impressive, the full realization of automatic repair in practical, large-scale software development environments remains a challenging goal that will require further advancements in AI technology and software engineering practices.

FULLY AUTOMATIC AI AGENT ARE IN EARLY STAGE

Despite the promising advancements, automatic code repair using AI face several significant challenges.

- *Program comprehension:* While Devin has demonstrated enhanced code context understanding compared to previous systems, it still encounters difficulties when faced with complex projects involving numerous interdependent components. For instance, in the case of scikit-learn-11542,[23] Devin identified only two instances with inconsistent default values. However, a comprehensive analysis starting from the RandomForestClassifier would reveal five such instances. This discrepancy highlights the need for more sophisticated algorithms capable of traversing and understanding complex dependency graphs in large-scale software projects.
- *Verification and testing:* Devin's current approach for verification and testing relies heavily on human-written test cases and brute force input generation techniques. While this methodology can be effective for simple problems, it falls short when dealing with the

CASE STUDY 1: UNDERSTANDING THE RANGE OF A VARIABLE.

I will provide a code that has bug: CWE-119…(shorten to save space)

```
static void  tokenadd( struct  jv_parser *p,   char  c) {
  assert(p->tokenpos <= p->tokenlen);
  /**<start-bug>**/
  if  (p->tokenpos == p->tokenlen) {
     /**<end-bug>**/
     p->tokenlen = p->tokenlen * 2 + 256;
     p->tokenbuf = jv_mem_realloc(p->tokenbuf, p->tokenlen);
  }
  assert(p->tokenpos < p->tokenlen);
  p->tokenbuf[p->tokenpos++] = c;
}
```

Reply from ChatGPT (shorten to save space): The issue here is that there's a potential to write outside of the allocated memory for *tokenbuf* in the last line of code.

```
…
/**<start-bug>**/
if(p->tokenpos >= p->tokenlen){
  p->tokenlen = p->tokenlen*2 + 256;
  char* newbuf = jv_mem_realloc(p->tokenbuf, p->tokenlen);
  if(newbuf == NULL) {
      fprintf(stderr, "Memory allocation failed\n");
      return ;
  }
  p->tokenbuf = newbuf;
}
/**<end-bug>**/
assert(p->tokenpos<p->tokenlen);
…
```

Additional knowledge: *Off-by-one error in the tokenadd function allows remote attackers to cause a denial of service (crash), which triggers a heap-based buffer overflow. If p-> tokenpos is more than or equal to the end of buffer, resize the buffer. The end of buffer should be p->* tokenlen -1 *not p->* tokenlen.

Reply from ChatGPT (shorten to save space):

```
…
/**<start-bug>**/
if(p->tokenpos>=p->tokenlen-1){
  p->tokenlen = p->tokenlen*2 + 256;
  char* newbuf = jv_mem_realloc(p->tokenbuf, p->tokenlen);
  if(newbuf == NULL) {
      fprintf(stderr, "Memory allocation failed\n");
      return ;
  }
   p->tokenbuf = newbuf;
}
/**<end-bug>**/
…
```

complexities of real-world software systems. For example, in scikit-learn-25744,[24] Devin correctly identified that the issue stemmed from the min_samples_split parameter and implemented an error message for integer values less than 2. However, it failed to verify the error condition for float values of min_samples_split, which should be constrained between 0.0 and 1.0 (exclusive) when representing a percentage. This oversight underscores the potential dangers of incomplete verification, particularly in critical systems where such oversights could lead to severe consequences.

- *Contextual understanding:* Automatic repair systems must not only fix the immediate bug but also ensure that the repair aligns with the broader context of the software, including design patterns, coding standards, and project-specific requirements. This level of contextual understanding remains a significant hurdle for current AI systems.

FUTURE DIRECTIONS IN AUTOMATIC CODE REPAIR

Recent research in AI-driven program repair has shown promising results, particularly in addressing well-defined programming tasks of limited scope.[25,26,27] These works have made significant strides by letting AI agents leverage static and dynamic analysis tools to examine compilation information and code output. This integrated approach guides AI agents in their repair efforts, improving the accuracy and reliability of the generated fixes. However, ensuring the correctness and reliability of AI-generated repairs remains a critical challenge, particularly as we move toward more complex systems. The field of AI-driven program repair continues to evolve, with several promising areas for future research.

Advanced program understanding

While recent models have improved in understanding code context, they still struggle with grasping the full scope of a program, including external dependencies, project-specific conventions, and broader architectural considerations. Developing more sophisticated techniques to capture semantic information and programmer intent is crucial for the future of AI-driven program repair. This may involve leveraging advanced natural language processing techniques to better interpret code comments and documentation.[28,29] Incorporating program dependency analysis could enhance the AI's understanding of the context and potential impact of repairs.[2] Additionally, utilizing machine-learning models trained on vast codebases could help in recognizing common patterns and idioms in software design.[4,6]

Rigorous verification and testing

One of the most significant challenges is ensuring the correctness of AI-generated patches. While AI models can generate plausible fixes, they may introduce new bugs or fail to fully address the underlying issue. Developing robust verification mechanisms for AI-generated patches remains an open problem. This involves integrating formal verification techniques with AI-generated repairs to provide mathematical guarantees of correctness.[25] Developing specialized testing frameworks that can automatically generate comprehensive test suites for AI-repaired code would help ensure the reliability of the fixes.[30] Additionally, utilizing symbolic execution and model checking techniques would allow for systematic exploration of the state space of repaired programs.[29]

Multilevel software reasoning

Enhancing AI models' ability to reason about software at various levels of abstraction is essential for comprehensive program repair. For example, GPT-o1 can reason through complex tasks and solve harder problems than previous models in science, coding, and math.[31] Future work could focus on developing hierarchical models that can simultaneously consider low-level code logic and high-level software system architectures.[32] Exploring reinforcement learning approaches might allow AI agents to learn from the consequences of their repair decisions across different abstraction levels.[33,34,35] By improving the AI's ability to reason at multiple levels, we can expect more sophisticated repairs that consider both local code improvements and their global impact on the system.

Explainability and transparency

From the perspective of AI agents, even as they are expected to work autonomously, the role of human supervision remains crucial, especially in the development and maintenance of critical systems. This underscores the importance of explainability and transparency in AI-driven program repair. Motivated by the need to bridge the gap between AI capabilities and human oversight, future work in this area should focus on several key aspects. Shapley additive explanation values could quantify the importance of different code features (for example, specific lines, functions, or dependencies) in the AI's decision to make a particular repair. This would allow human supervisors to understand which parts of the code most influenced the AI's choice of repair strategy.[36] Developing sophisticated attention mechanisms could highlight specific parts of the code that influence the AI's repair decisions, providing insight into the agent's focus and reasoning process.[37]

By addressing these key areas, researchers aim to bridge the gap between current capabilities and the vision of fully autonomous AI agents capable of general-purpose program repair. While this goal remains distant, ongoing advancements in these areas continue to push the boundaries of what's possible in AI-driven software development and bug fixing.

Our investigation into the capabilities of LLMs in program repair reveals a nuanced landscape with significant implications for software engineering. In the realm of human–LLM collaboration, our findings demonstrate a promising synergy, where human expertise in providing contextual knowledge significantly enhances LLMs' effectiveness in bug repair processes. This collaborative approach shows great potential for improving software development and bug fixing practices, particularly in complex systems. However, the results for autonomous AI agent repair indicate that we are still far from achieving reliable, fully autonomous code repair using LLMs alone. These findings lead us to conclude that while LLMs represent a powerful tool in software engineering, they are not yet ready to replace human expertise in program repair. The most promising path forward appears to be a hybrid approach that leverages the strengths of both human developers and LLMs. As we move forward, it is crucial to focus on enhancing LLMs' contextual understanding, developing more sophisticated human–LLM interfaces, and improving LLMs' ability to reason about and verify their proposed solutions. By maintaining a balanced perspective and working toward solutions that harmoniously combine human expertise and AI, we can continue to advance the field of software development.

ACKNOWLEDGMENT

Peng Liu was supported by National Science Foundation (NSF) CNS-2019340, NSF ECCS-2140175, and National Institute of Standards and Technology 60NANB22D144. Xiaoyan Sun is supported by NSF DGE2105801.

REFERENCES

1. Z. Wang, L. Zhang, and P. Liu, "ChatGPT for software security: Exploring the strengths and limitations of ChatGPT in the security applications," 2023, *arXiv:2307.12488*.
2. L. Zhang, Q. Zou, A. Singhal, X. Sun, and P. Liu, "Evaluating large language models for real-world vulnerability repair in C/C++ code," in *Proc. 10th ACM Int. Workshop Secur. Privacy Analytics*, New York, NY, USA, 2024, pp. 49–58, doi: 10.1145/3643651.3659892.
3. H. Joshi, J. C. Sanchez, S. Gulwani, V. Le, I. Radiček, and G. Verbruggen, "Repair is nearly generation: Multilingual program repair with LLMs," in *Proc. 37th AAAI Conf. Artif. Intell. (AAAI)*, 2023, vol. 37, no. 4, pp. 5131–5140, doi: 10.1609/aaai.v37i4.25642.
4. S. Rozendal. "Breakthrough in AI agents?" Effective Altruism Forum. Accessed: Apr. 10, 2024. [Online]. Available: https://forum.effectivealtruism.org/posts/3brE2Mt6qC-72cQvzL/breakthrough-in-ai-agents-on-devin-the-zvi-linkpost
5. A. Sellen, and E. Horvitz. "The rise of the AI co-pilot: Lessons for design from aviation and beyond," 2023, *arXiv:2311.14713*.
6. Y. Bajpai et al., "Let's fix this together: Conversational debugging with GitHub Copilot," in *Proc. IEEE Symp. Vis. Lang. Human-Centric Comput. (VL/HCC)*, 2024, pp. 1–12.
7. B. Yang et al., "CREF: An LLM-based conversational software repair framework for programming tutors," 2024, *arXiv:2406.13972*.
8. H. Pearce, B. Tan, B. Ahmad, R. Karri, and B. Dolan-Gavitt, "Examining zero-shot vulnerability repair with large language models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, 2023, pp. 2339–2356, doi: 10.1109/SP46215.2023.10179420.

9. U. Kulsum, H. Zhu, B. Xu, and M. d'Amorim, "A case study of LLM for automated vulnerability repair: Assessing impact of reasoning and patch validation feedback," in *Proc. 1st ACM Int. Conf. AI-Powered Softw.*, New York, NY, USA, 2024. pp. 103–111, doi: 10.1145/3664646.3664770.
10. S. Kang, J. Yoon, and S. Yoo, "Large language models are few-shot testers: Exploring LLM-based general bug reproduction," 2022, *arXiv:2209.11515*.
11. C. S. Xia, and L. Zhang, "Keep the conversation going: Fixing 162 out of 337 bugs for $0.42 each using ChatGPT," 2023, *arXiv:2304.00385*.
12. N. Wadhwa et al., "Frustrated with code quality issues? LLMs can Help!" 2023, *arXiv:2309.12938*.
13. C. Lemieux, J. Priya Inala, S. K. Lahiri, and S. Sen, "CodaMosa: Escaping coverage plateaus in test generation with pre-trained large language models," in *Proc. IEEE/ACM 45th Int. Conf. Softw. Eng. (ICSE)*, Melbourne, Australia, 2023, pp. 919–931, doi: 10.1109/ICSE48619.2023.00085.
14. J. Cao, M. Li, M. Wen, and S-c Cheung, "A study on prompt design, advantages and limitations of ChatGPT for deep learning program repair," 2023, *arXiv:2304.08191*.
15. N. Parasaram et al., "The fact selection problem in LLM-based program repair," 2024, *arXiv:2404.05520*.
16. C. E. Jimenez et al., "SWE-bench: Can language models resolve real-world GitHub issues?" in *Proc. 12th Int. Conf. Learn. Representations*, 2024.
17. "Windsurf editor and Codeium extensions." Codeium. [Online]. Available: https://codeium.com/
18. "Engineering genius." Anysphere. [Online]. Available: https://anysphere.inc/
19. "Magic." Magic.dev. [Online]. Available: https://magic.dev/
20. "Replit – Build apps and sites with AI." Replit. [Online]. Available: https://replit.com/
21. "Cody - Business AI employee trained on your knowledge base." Cody AI. [Online]. Available: https://meetcody.ai/
22. C. E. Jimenez et al., "SWE-bench: Can language models resolve real-world GitHub issues?" 2024, *arXiv:2310.06770*.
23. "CognitionAI/devin-swebench -results." GitHub. [Online]. Available: https://github.com/CognitionAI/devin-swebench-results/blob/main/output_diffs/fail/scikit-learn__scikit-learn-11542-diff.txt
24. "CognitionAI/devin-swebench-results." GitHub. [Online]. Available: https://github.com/CognitionAI/devin-swebench-results/blob/main/output_diffs/fail/scikit-learn__scikit-learn-25744-diff.txt
25. Z. Fan, X. Gao, M. Mirchev, A. Roychoudhury, and S. H. Tan, "Automated repair of programs from large language models," 2022, *arXiv:2205.10583*.
26. I. Bouzenia, P. Devanbu, and M. Pradel, "RepairAgent: An autonomous, LLM-based agent for program repair," 2024, *arXiv:2403.17134*.
27. C. Qian et al., "Iterative experience refinement of software-developing agents," 2024, *arXiv:2405.04219*.
28. Y. Shinyama, Y. Arahori, and K. Gondow, "Analyzing code comments to boost program comprehension," in *Proc. 25th Asia-Pacific Softw. Eng. Conf. (APSEC)*, 2018, pp. 325–334, doi: 10.1109/APSEC.2018.00047.
29. W. Zhou, L. Zhang, L. Guan, P. Liu, and Y. Zhang, "What your firmware tells you is not how you should emulate it: A specification-guided approach for firmware emulation," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, 2022, pp. 3269–3283.
30. C. Oliveira, A. Aleti, L. Grunske, and K. Smith-Miles, "Mapping the effectiveness of automated test suite generation techniques," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 771–785, Sep. 2018, doi: 10.1109/TR.2018.2832072.
31. "Introducing OpenAI o1-preview." OpenAI. [Online]. Available: https://openai.com/index/introducing-openai-o1-preview/
32. H. Muccini, P. Inverardi, and A. Bertolino, "Using software architecture for code testing," *IEEE Trans. Softw. Eng.*, vol. 30, no. 3, pp. 160–171, Mar. 2004, doi: 10.1109/TSE.2004.1271170.
33. L. Zhang, P. Liu, Y.-H. Choi, and P. Chen, "Semantics-preserving reinforcement learning attack against graph neural networks for malware detection," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 2, pp. 1390–1402, Mar./Apr. 2022, doi: 10.1109/TDSC.2022.3153844.
34. N. T. Islam, J. Khoury, A. Seong, G. D. L. T. Parra, E. Bou-Harb, and P. Najafirad, "LLM-powered code vulnerability repair with reinforcement learning and semantic reward," 2024, *arXiv:2401.03374*.
35. R. Gupta, A. Kanade, and S. Shevade, "Deep reinforcement learning for syntactic error repair in student programs," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 930–937, doi: 10.1609/aaai.v33i01.3301930.

36. Q. Zou, L. Zhang, A. Singhal, X. Sun, and P. Liu, “Using explainable AI for neural network-based network attack detection,” *Computer*, vol. 57, no. 5, 2024, pp. 78–85, doi: 10.1109/MC.2023.3342602.
37. Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

LAN ZHANG is an assistant professor at the School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86005 USA. Contact her at lan.zhang@nau.ed.

ANOOP SINGHAL is a senior computer scientist in the Computer Security Division at the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA. Contact him at anoop.singhal@nist.gov.

QINGTIAN ZOU is a postdoctoral researcher at the University of Texas, Southwestern Medical Center, Dallas, TX 75390 USA. Contact him at qingtian.mill.zou@gmail.com.

XIAOYAN SUN is an associate professor with Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609 USA. Contact her at xsun7@wpi.edu.

PENG LIU is the Raymond G. Tronzo, MD Professor of Cybersecurity and serves as the director of the Cyber Security Lab at The Pennsylvania State University, State College, PA 16803 USA. Contact her at pxl20@psu.edu.

IEEE COMPUTER SOCIETY

www.computer.org

PURPOSE: Engaging professionals from all areas of computing, the IEEE Computer Society sets the standard for education and engagement that fuels global technological advancement. Through conferences, publications, and programs, IEEE CS empowers, guides, and shapes the future of its members, and the greater industry, enabling new opportunities to better serve our world.

OMBUDSMAN: Contact ombudsman@computer.org.

CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

PUBLICATIONS AND ACTIVITIES

***Computer*:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

Periodicals: The IEEE CS publishes 12 magazines, 18 journals

Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.

Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.

Technical Communities: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

Conferences/Education: The IEEE CS holds more than 215 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The IEEE CS offers three software developer credentials.

AVAILABLE INFORMATION

To check membership status, report an address change, or obtain information, contact help@computer.org.

IEEE COMPUTER SOCIETY OFFICES

WASHINGTON, D.C.:
2001 L St., Ste. 700,
Washington, D.C. 20036-4928

Phone: +1 202 371 0101

Fax: +1 202 728 9614

Email: help@computer.org

LOS ALAMITOS:
10662 Los Vaqueros Cir.,
Los Alamitos, CA 90720

Phone: +1 714 821 8380

Email: help@computer.org

IEEE CS EXECUTIVE STAFF

Interim Executive Director: Anne Marie Kelly
Director, Governance & Associate Executive Director: Anne Marie Kelly
Director, Conference Operations: Silvia Ceballos
Director, Information Technology & Services: Sumit Kacker
Director, Marketing & Sales: Michelle Tubb
Director, Membership Development: Eric Berkowitz
Director, Publications & Special Projects: Robin Baldwin

IEEE CS EXECUTIVE COMMITTEE

President: Hironori Washizaki
President-Elect: Grace A. Lewis
Past President: Jyotika Athavale
Vice President: Nils Aschenbruck
Secretary: Yoshiko Yasuda
Treasurer: Darren Galpin
VP, Member & Geographic Activities: Andrew Seely
VP, Professional & Educational Activities: Cyril Onwubiko
VP, Publications: Charles (Chuck) Hansen
VP, Standards Activities: Edward Au
VP, Technical & Conference Activities: Terry Benzel
2025–2026 IEEE Division VIII Director: Cecilia Metra
2024–2025 IEEE Division V Director: Christina M. Schober
2025 IEEE Division V Director-Elect: Leila De Floriani

IEEE CS BOARD OF GOVERNORS

Term Expiring 2025:
İlkay Altintaş, Joaquim Jorge, Rick Kazman, Carolyn McGregor, Andrew Seely

Term Expiring 2026:
Megha Ben, Terry Benzel, Mrinal Karvir, Andreas Reinhardt, Deborah Silver, Yoshiko Yasuda

Term Expiring 2027:
Sven Dickinson, Alfredo Goldman, Daniel S. Katz, Yuhong Liu, Ladan Tahvildari, Damla Turgut

IEEE EXECUTIVE STAFF

Executive Director and COO: Sophia Muirhead
General Counsel and Chief Compliance Officer: Ahsaki Benion
Chief Human Resources Officer: Cheri N. Collins Wideman
Managing Director, IEEE-USA: Russell Harrison
Chief Marketing Officer: Jayne O'Brien
Chief Publication Officer and Managing Director: Steven Heffner
Staff Executive, Corporate Activities: Donna Hourican
Managing Director, Member and Geographic Activities: Cecelia Jankowski
Chief of Staff to the Executive Director: Allison Taylor
Managing Director, Educational Activities: Jamie Moesch
IEEE Standards Association Managing Director: Alpesh Shah
Chief Financial Officer: Kelly Armstrong
Chief Information Digital Officer: Jeff Strohschein
Managing Director, Conferences, Events, and Experiences: Marie Hunter
IEEE Standards Association Managing Director: Alpesh Shah
Chief Governance Officer: Donna Hourican

IEEE OFFICERS

President & CEO: Kathleen A. Kramer
President-Elect: Mary Ellen Randall
Past President: Thomas M. Coughlin
Director & Secretary: Forrest D. Wright
Director & Treasurer: Gerardo Barbosa
Director & VP, Publication Services & Products: W. Clem Karl
Director & VP, Educational Activities: Timothy P. Kurzweg
Director & VP, Membership and Geographic Activities: Antonio Luque
Director & President, Standards Association: Gary R. Hoffman
Director & VP, Technical Activities: Dalma Novak
Director & President, IEEE-USA: Timothy T. Lee

IEEE

revised 17 December 2025

EDITORS: Ciera Jaspan, Google, ciera@google.com
Collin Green, Google, colling@google.com

This article originally appeared in Software vol. 42, no. 3, 2025

DEPARTMENT: DEVELOPER PRODUCTIVITY FOR HUMANS

Software Development Is a Team Sport

Claire Taylor, Marie Huber, Qiao Ma, Rayven Plaza, Alison Chang, and Jie Chen

A well-coordinated, high-functioning team is critical to developing and delivering quality product experiences at a competitive pace. However, the field of software engineering has historically emphasized individuals when studying outcomes like productivity and velocity. Understanding these outcomes at a team or product level requires more than simply aggregating individual-level measures: Teamwork is necessary to avoid counterproductive individual efforts, and collaboration with nondeveloper roles (e.g., user experience, product) contribute meaningfully to these higher-order outcomes.

To understand how individual efforts combine to determine group outcomes like productivity, we need to understand how engineers and other roles work together when creating software products. In this column, we present research undertaken to identify collaboration through observable events and assess team functioning through self-report surveys; together, these help us understand software development as a team sport.

YOU WORK WITH WHO?

Collaboration plays a critical role in the way teams get their work done, but it's difficult to define specifics: Who works with whom? When and how do they interact? What patterns exist across the organization? At a large, global company like Google, the answers to these questions are continuously shifting, as projects begin and end, individuals change roles or join the company, and organizational priorities evolve. We already had existing data sources like management chain (e.g., who reports to the same manager) and manually maintained team assignments (e.g., who works on the same product or project) that indicated possible collaborators, but both methods assume that everyone on a given team works together (which we knew wasn't true at Google) and these collaborations are stable over time and across individuals (although research shows they are fluid and engineers may not agree about team composition[1]). These groupings don't fully account for how employees work across nominal teams and job functions. This knowledge motivated us to develop a method for measuring collaboration that could describe what collaborations look like yesterday, today, and tomorrow for each engineer, creating a flexible way to determine who comprised an engineer's teammates. (We focused on measuring engineers' collaborations with other employees to narrow the problem space. We have plans to extend the metric to all Googlers in the future.)

MEASURING COLLABORATION

In reviewing the literature, we saw that surveys,[2] interviews,[3] and other user-centered data provided rich information about collaborations, but they weren't scalable over time. In contrast, logs-based approaches enabled ongoing measurement,[4,5] which encouraged us to use the signals we were already collecting from a number of tools commonly used for collaborative work at Google.[6] We understood there were some collaborative interactions—like impromptu in-person conversations with no associated logged events—that would be blind spots for our metric, but decided the tradeoffs were reasonable and that collaborators engaging in these behaviors would have other collaborative events that would be logged.

Digital Object Identifier 10.1109/MS.2025.3539403
Date of current version: 11 April 2025

We operationalized collaboration as "any interaction between employees where those involved are aware of one another's actions, as observed through logs within a predefined list of coding and communication tools." Using this definition, we proposed an initial version of our metric that used logs to identify collaborative events between employees that took place in each tool (e.g., Meet, Docs, Sheets, Slides, Chat, code review tooling, bug tracker). We summed events between each engineer and their colleagues for each tool, and then applied a weight to account for differences between tools (e.g., colleagues may send many chats within a few minutes, but may only meet a few times per week or month). We then looked across these per tool values to identify the most common collaborators for each engineer. We applied the notion of "meaningful interactions"[5] by excluding potential collaborative interactions that were too large or noisy (e.g., large group meetings) to be reliable signals of collaboration, and that occurred far apart in time (e.g., document comments that happen months apart).

We wanted to understand who engineers viewed as their collaborators to validate our approach and refine the metric. We conducted a short survey, asking engineers to review their list of top 13 collaborators identified by the metric and write in any collaborators it missed. The metric demonstrated high accuracy (both precision and recall) in identifying the top collaborators. Metric-identified collaborators more closely matched who engineers viewed as their collaborators, compared to those inferred through management chain or team assignments. We made a few small refinements to improve metric performance, namely requiring use of two or more tools by collaborators, adjusting weights for each tool, and trimming the long tail of collaborators for each engineer if they comprised less than 1% of their collaborative activities.

In addition to showing high accuracy, our refined metric demonstrated high-level patterns that matched with prior qualitative insights and common assumptions within the company. More senior engineers tend to collaborate more often and with more people, consistent with their job descriptions and expectations, whereas more junior engineers collaborate less frequently and have fewer collaborators. Transferring roles and teams within the company impacts collaboration, with recently transferred engineers collaborating less. Collaborative tool use varied across regions, most notably greater reliance on asynchronous collaboration tools by engineers outside of the Americas, where Google is largely concentrated.

CHARACTERIZING A HIGH-FUNCTIONING TEAM

Our new metric enabled us to know when collaboration was happening, but it didn't provide us with a notion of how that teamwork was going: good, bad, or otherwise. To measure the quality of collaborations, we began work to assess team functioning. In the context of software engineering and product development at Google, this means teammates across roles (e.g., engineering, product, user experience) collaborate effectively and efficiently to maximize velocity of software and product development.

WE WANTED TO UNDERSTAND WHO ENGINEERS VIEWED AS THEIR COLLABORATORS TO VALIDATE OUR APPROACH AND REFINE THE METRIC.

The quality of collaboration is not easily measured with logs, so we set out to develop a reliable survey instrument to triangulate with other data and provide a complete picture of team dynamics, including what helps teams thrive or holds them back. We experimented with adapting existing survey tools for measuring team processes,[7] but found these did not capture the nuances of software and product development or Google culture.

Both internal[8] and external[7] research shows that higher-functioning teams are more productive and produce higher-quality outputs. Teams at Google sometimes struggle with collaborative and operational processes that impact team functioning and outcomes like productivity. Teammates can have differing or conflicting expectations about how work will get done. Variance in skills and expertise across roles can exacerbate these differences. This knowledge shaped our goals for developing a valid and reliable measure of team functioning.

- We wanted to understand how team functioning fit into the larger context of software and product development at Google. What drives higher (and lower) team functioning? Are some types of teams or contexts more likely to demonstrate higher functioning? How does team functioning relate to outcomes like productivity, quality, and velocity? Developing a reliable measure would enable us to use it alongside our portfolio of logs- and survey-base metrics.
- We wanted to empower teams and leaders with a useful tool that would help them identify bottlenecks and pain points that might be holding their teams back. Diagnosing problems could help focus teams on working together to find solutions and help leaders and organizations prioritize interventions and resources to address needs to accelerate desired outcomes.

To understand the experience of team functioning, we needed to engage directly with engineers and nonengineers within the company. They were best positioned to help us understand how they thought about the different elements of team functioning, what those looked like on their team, and how they talked about them in their day-to-day work. In addition to the quarterly large-scale survey run with Google engineers (called *EngSat*),[9] we run annual surveys with smaller nonengineering functions (like product managers and UX designers and researchers) that work on software product development teams. We created a survey scale to measure team functioning that we could use across this family of satisfaction surveys (collectively called the *Sat Surveys*), enabling us to collect the responses we needed to build the company-wide picture we were aiming for.

MOVING FROM DEFINITION TO MEASUREMENT

Developing Potential Survey Items

To get measurement right, we followed established best practices from the field of psychometrics to guide development of a survey-based multi-item scale that would allow for employees to self-report about their team's functioning.[10] Multi-item (e.g., multiple survey questions) measures are better suited to complex and multidimensional concepts or latent constructs (like team functioning), offering greater reliability and sensitivity for triangulation, predictive validity, and detection of differences in smaller samples. Single-item measures are better suited for simple, unidimensional concepts.

We started with a comprehensive literature review to define the domain and generate potential survey items. Next, we asked subject matter experts representing different departments and roles across Google to review our proposed set of items, helping to ensure we covered the full domain of team functioning without gaps or redundancy and that our items were of high quality. After incorporating this feedback, we engaged one-on-one with Googlers representing the engineering and nonengineering roles, representing the intended audience for the measure. These interviews complemented the subject matter expert reviews: We asked Googlers if the questions made sense to them and felt relevant to their work and team. Incorporating this round of feedback meant we had a great set of potential survey items, but far too many (almost 100 items!) for inclusion in a typical survey, and more than we practically needed to measure team functioning.

Refining the Survey Measure

The next steps in scale development focus on reducing the number of items and determining how those remaining items are related. Determining which items assess related or overlapping concepts ensures important concepts are covered by the scale without redundancy. Determining this functional and parsimonious set of items requires a larger dataset.[10] We recruited 245 engineers and nonengineers from product development teams within Google to participate in 30-min moderated group sessions. Researchers briefly set context (~5 min), so participants could use most of the session to complete survey items.

Our analysis indicated that four factors accounted for the majority of measurement variance, suggesting this was likely the correct number subscales for our overall measure. We conducted a factor analysis with a four-factor solution that enabled removal of poor

TABLE 1. Team functioning factor definitions and sample items.

Team functioning factor	Definition	Sample items
Team processes and visibility	Team uses tools and processes that add value without burden, that supports visibility within the team, effective information management, and timely and efficient flow of information within the team.	Aligns with one another on the tools and technologies we use for shared workflows?
		Uses shared team processes and tools in ways that are helpful?
Team culture	Team maintains a safe, respectful, and collaborative team environment that encourages constructive discussion and learning. Team members can make high-value contributions aligned with their roles, that are recognized and valued by others within the team.	Creates a safe environment for sharing mistakes and lessons learned?
		Enables all team members to raise topics for discussion or feedback with the team?
Strategic alignment	Team ensures the work they do is aligned with broader organizational priorities, user needs and priorities, and high-value business goals.	Makes sure our work ladders up to broader organizational goals and priorities?
		Has a strong understanding within the team of the value our product provides to users?
Balanced team workload	Team takes on a manageable amount of work, and effectively plans for and executes it in a way that considers the capacity of each team member. Team members are able to deliver quality work while maintaining their well-being.	Takes on the right amount of work as a team?
		Aims to work at a pace that prioritizes wellbeing and quality work across the team?

All items use the question stem: "In the cross-functional team you interact with day to day on the products you work on, how well do you feel that the team…".

performing items, narrowing the item pool to 57. We conducted a second factor analysis on the 57 items, again extracting four factors. Items with the strongest association with each factor were selected to form distinct subscales of the overall team functioning measure, described in Table 1.

These factors mirrored areas identified in our literature review, excepting "strategic alignment," which captures elements critical to product development in a large organization that may not generalize to teams operating in different contexts. We further refined the subscales by evaluating each item based on its factor loading, theoretical relevance, item-total correlation, and contribution to Cronbach's α. This resulted in concise versions of each subscale, optimal for use in the Sat Surveys.

Teamwork Drives Outcomes

The team functioning scale and its subscales demonstrated meaningful, positive relationships with important self-reported outcomes measured on the Sat Surveys. Specifically, Googlers who reported higher levels of team functioning reported feeling more productive at work and greater satisfaction with their role at Google. They were also more satisfied with their involvement across the stages of product development, and quality and speed at which their team ships products to users (e.g., team velocity). These relationships held when accounting for other characteristics (e.g., job level, tenure) and across engineering and nonengineering roles, demonstrating that high-functioning teams play an important role in the success of both individuals and their teams.

We also investigated the relationships between team functioning and logs-based measures of productivity for engineers focused on code writing and reviewing, but these relationships were more nuanced. Team functioning was not a strong predictor of these measures on its own, although the relationship strengthened when we accounted for job level and manager/individual contributor status. This suggests a more complex relationship that may be mediated or cofounded by these or other factors, and we plan to continue our investigation.

As this work continues, identifying what differentiates higher and lower functioning teams, including patterns of collaboration or team-level characteristics, may help guide teams toward improvement and enable the team functioning scale to be used as a diagnostic tool for teams. We're working to create a guide for scale administration and result calculation, so teams can interpret their results and facilitate a dialogue around their strengths and challenges. This can make the information more useful for teams, especially when paired with guidance for next steps. The ultimate goal is to connect teams with tools and resources to support their growth, and identify spaces where additional company support to individuals or teams may lead to improved outcomes.

Our findings demonstrate the value of creating and maintaining team processes that support the group's work as a counterbalance to focus on individual productivity and outcomes. Working to understand and improve team functioning, not just individual performance, drives better software development outcomes.

REFERENCES

1. A. Ruvimova and G. Rasch et al., "An exploratory study of productivity perceptions in software teams," in *Proc. IEEE/ACM 44th Int. Conf. Softw Eng (ICSE)*, Pittsburgh, PA, USA, 2022, pp. 99–111, doi: 10.1145/3510003.3510081.
2. I. Boughzala and G. J. De Vreede, "Evaluating team collaboration quality: The development and field application of a collaboration maturity model," *J. Manage. Inf. Syst.*, vol. 32, no. 3, pp. 129–157, Jul. 2015, doi: 10.1080/07421222.2015.1095042.
3. F. Batarseh, J. Usher, and J. Daspit, "Collaboration capability in virtual teams: Examining the influence on diversity and innovation," *Int. J. Innov. Manage.*, vol. 21, no. 4, May 2017, Art.no. 1750034, doi: 10.1142/S1363919617500347.
4. E. DeFilippis et al., "Collaborating during Coronavirus: The impact of COVID-19 on the nature of work," National Bureau of Economic Research, Cambridge, MA, USA, Jul. 2020. [Online]. Available: https://doi.org/10.3386/w27612
5. L. Yang et al., "The effects of remote work on collaboration among information workers," *Nature Hum. Behav.*, vol. 6, no. 1, pp. 43–54, Jan. 2022, doi: 10.1038/s41562-021-01196-4.
6. C. Jaspan et al., "Enabling the study of software development behavior with cross-tool logs," *IEEE Softw.*, vol. 37, no. 6, pp. 44–51, Nov./Dec. 2020, doi: 10.1109/MS.2020.3014573.
7. J. Mathieu et al., "The development and construct validity of a team processes survey measure," *Org. Res. Meth.*, vol. 23, no. 3, pp. 399–431, Apr. 2019, doi: 10.1177/1094428119840801.
8. "Understand team effectiveness." Google re:Work. Accessed: Dec. 17, 2024. [Online]. Available: https://rework.withgoogle.com/en/guides/understanding-team-effectiveness
9. S. D'Angelo et al., "Measuring developer experience with a longitudinal survey," *IEEE Softw.*, vol. 41, no. 4, pp. 19–24, Jul./Aug. 2024, doi: 10.1109/MS.2024.3386027.
10. G. Boateng et al., "Best practices for developing and validating scales for health, social, and behavioral research: A primer," *Front. Public Health*, vol. 6, Jun. 2018, Art.no. 149, doi: 10.3389/fpubh.2018.00149.

CLAIRE TAYLOR is a user experience researcher on the Engineering Productivity Research team, Google, Seattle, WA 98109 USA. Contact her at clairetaylor@google.com.

MARIE HUBER is a user experience researcher on the TechIntel team at Google in San Francisco, CA 94105 USA. Contact her at mariehuber@google.com.

QIAO MA is a quantitative user experience researcher on the TechIntel team at Google, Mountain View, CA 94043 USA. Contact her at qiaoma@google.com.

RAYVEN PLAZA is a quantitative user experience researcher on the Engineering Productivity Research team at Google in New York, NY 10011 USA. Contact her at rayven@google.com.

ALISON CHANG is a software engineer on the Engineering Productivity Research team at Google at Mountain View, CA 94043 USA. Contact her at alisonchang@google.com.

JIE CHEN is a data scientist on the Developer Data Science team at Google in New York, NY 10011 USA. Contact her at jennjiechen@google.com.

EDITOR: Nir Kshetri, The University of North Carolina at Greensboro, nbkshetr@uncg.edu

DEPARTMENT: COMPUTING'S ECONOMICS

This article originally appeared in Computer vol. 58, no. 9, 2025

From Crisis to Comeback: The Evolution and Resurgence of Blockchain Post-2022 Crypto Winter

Nir Kshetri, *The University of North Carolina at Greensboro*

This article explores how blockchain technology is recovering and adapting following the 2022 crypto market downturn. It highlights major challenges faced, regulatory changes, and new developments shaping its renewed adoption in various industries.

Blockchain was conceived as a "trust machine" to counter centralized system failures, yet its promise of secure, decentralized transactions has been repeatedly undermined by industry-wide trust issues. Recent "black swan events" have further eroded market confidence.[1] One such black swan event was the 2022 collapse of cryptocurrency exchange platform FTX, which filed for bankruptcy in November with an estimated US$32 billion valuation.[2] The FTX scandal not only failed to shift the views of staunch blockchain skeptics but also alienated segments of the public who were previously neutral or open to crypto.[3]

A number of indicators suggest that the broader blockchain ecosystem—and the crypto market as one of its key components—is steadily recovering from the 2022 crypto winter. As of 5 June 2025, Bitcoin is trading at approximately US$104,657, slightly below its all-time high closing price of US$111,970 recorded on 22 May 2025. This marks a dramatic rebound from its lowest point in 2022, when it closed at US$15,787.28 on 21 November, reflecting a 64.3% drop for that year.

While cryptocurrency prices often dominate headlines, the broader blockchain landscape has been evolving in parallel, with significant innovations tailored for enterprise and institutional applications. With a focus on speed and sustainability, modern blockchain platforms were created to overcome the limitations of early systems like Bitcoin and support real-world business operations. Enterprise adoption of blockchain technologies is gaining momentum, particularly in scenarios necessitating inter-organizational data exchange. Advanced use cases now include trade finance, identity management, and enterprise resource planning functions, such as supply chain oversight. Concurrently, decentralized finance is driving increased platform demand by offering alternative business frameworks that challenge conventional financial systems.[4]

MODERN BLOCKCHAIN PLATFORMS WERE CREATED TO OVERCOME THE LIMITATIONS OF EARLY SYSTEMS LIKE BITCOIN AND SUPPORT REAL-WORLD BUSINESS OPERATIONS.

This technological and market momentum is beginning to translate into renewed investor confidence, as signs point to a gradual rebound in venture funding for crypto and blockchain startups. In 2023, venture capital investment in crypto and blockchain startups fell sharply—down 68% from 2022 levels. Notably, most of the 2022 funding occurred in the first half of the year, before a series of major crypto firm collapses triggered a steep decline in the second half. The significant drop in 2023 was largely anticipated, driven by a combination of macroeconomic pressures, regulatory uncertainty, and lingering fallout from

Digital Object Identifier 10.1109/MC.2025.3578849
Date of current version: 27 August 2025

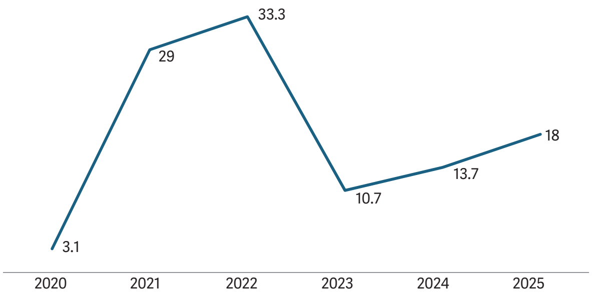


FIGURE 1. Venture capital in crypto and blockchain startups (US$, billion). (Data sources: For 2020, Melinek[6]; for 2021–2024, Khatri[7]; for 2025, Binance News.[8])

earlier industry failures.[5] However, data from 2024 and early 2025 indicate a positive shift, with venture funding steadily increasing as investors regain confidence amid clearer regulatory signals and continued technological advancements in the sector (Figure 1).

This article examines the evolution of blockchain technology from its initial promise as a decentralized "trust machine" to its current role in restoring confidence after the 2022 crypto winter. It analyzes key industry failures, regulatory shifts, and emerging innovations that are driving blockchain's resurgence across enterprise, finance, and decentralized infrastructure sectors.

THE 2022 CRYPTO WINTER

The cryptocurrency market is known for its extreme volatility, marked by cycles of rapid growth followed by sharp declines. Understanding such downturns—often referred to as "crypto winters"—is important because they highlight the structural vulnerabilities of the digital asset ecosystem and the factors that can erode investor confidence and market stability.[9] The 2022 crypto winter was partly triggered by high U.S. inflation and aggressive Federal Reserve interest rate hikes. Key contributors included the May collapse of Luna and TerraUSD, which dragged Bitcoin to its lowest level since 2020, the crash of US$10 billion crypto hedge fund Three Arrows Capital due to failed coin investments, and FTX's November bankruptcy after mismanaging customer funds.[10]

Table 1 presents a comparative overview of four major collapses in the cryptocurrency industry—Terra/Luna, Celsius Network, FTX, and Voyager Digital. Each case highlights distinct triggers, such as algorithmic instability, risky lending practices, fraudulent activities, and misleading consumer assurances. Together, these collapses underscore the vulnerabilities of loosely regulated digital asset markets and the cascading effects of interlinked crypto platforms. The timeline and aftermath of each incident offer critical insights into the systemic risks and governance failures that continue to shape the crypto ecosystem.

By 21 December 2022, the global crypto market cap had plunged to US$845 billion—a 65% drop from the previous year. Bitcoin and Ether saw sharp declines, falling from their 2021 highs of US$69,000 and US$4,800 to US$17,000 and US$1,200, respectively. This downturn marked the 2022 crypto winter, driven by falling prices, low trading volumes, and waning investor confidence. The collapse of TerraUSD and Luna in 2022 May triggered the crisis, which deepened after FTX's bankruptcy and fraud allegations, leading to further fallout like BlockFi's collapse.[9] After crypto's previous two-year hibernation ended in 2020, the sector surged to nearly US$3 trillion in total assets by November 2021, before crashing below US$1 trillion by June 2022.[13] Fears of tighter regulation and broader economic pressures, including inflation and recession concerns, further chilled the market.[9]

The 2022 crypto winter underscored the urgent need for regulatory clarity, strong governance, and robust investor protections to support the sector's long-term resilience and growth.[9] Regulations announced in spring 2021 were revised after Terra's collapse, with some jurisdictions drafting new rules to mitigate the systemic risks posed by failed stablecoin systems.[13]

TABLE 1. Major crypto collapses: causes, timelines, and consequences.

Company	Collapse timeline	Cause of collapse	Aftermath/consequence
FTX/Sam Bankman-Fried[2]	November 2022–March 2024	– Misuse of customer funds (FTX → Alameda) – Corporate control failures – Binance withdrawal – Lack of liquidity	– 25-year prison sentence for Sam Bankman-Fried – Chapter 11 bankruptcy – Billions lost – Potential customer repayment efforts underway
Terra/Do Kwon[11]	May 2022	– Algorithmic stablecoin failure (UST) – Massive market crash – Poor risk management	– Terra ecosystem collapse – Do Kwon arrested and charged – Severe loss for investors
Celsius Network[12]	June 2022–July 2022	– Mismanagement of customer deposits – Liquidity crisis – Failed business model	– Bankruptcy filing – Customers locked out of funds – Regulatory investigations
Voyager Digital[12]	July 2022	– Customer default on US$650 million loan – False claims about the U.S. Federal Deposit Insurance Corporation (FDIC) deposit insurance – Liquidity crisis	– Bankruptcy filing – FDIC investigation launched – Assets bought by FTX then put up for bid again after FTX bankruptcy – Binance.U.S. acquisition announced for US$1.02 billion

After announcing plans in March 2022 to regulate stablecoins under electronic payment laws, the U.K. government proposed in late May 2022 a separate set of amendments aimed at managing the collapse of stablecoin firms that could threaten systemic stability.[14] In mid-May 2022, following the collapse of Terra's UST and LUNA—which affected an estimated 280,000 South Koreans—financial authorities began considering stricter oversight of crypto exchanges. The issue was also discussed at a National Assembly emergency seminar focused on the crisis.[15]

CRYPTO MARKET RECOVERY AMID REGULATORY SHIFTS AND POLITICAL CHANGE

While it's unclear whether the crypto winter is definitively over, there are signs of meaningful improvement. Ongoing regulatory uncertainty and macroeconomic headwinds continue to weigh on the industry, yet indicators, such as higher transaction fees, increased developer activity, and a surge in smart contract deployment, suggest that underlying fundamentals are strengthening. Most analysts agree that the market is on a recovery path, even if a full-fledged bull run has yet to materialize.[16]

One argument suggests that the surge in cryptocurrency markets in the United States in late 2024 can be attributed to Donald Trump's victory in the U.S. presidential election. His pro-crypto stance—highlighted by promises to make the United States the "crypto capital," appoint a crypto-friendly Securities and Exchange Commission (SEC) chair, block a Federal Reserve-issued digital currency, and create a pro-crypto advisory council—renewed investor confidence. This political shift fueled broader market optimism and contributed to a sharp rally across major digital assets.[17] Critics have also pointed out that under SEC Chair Gary Gensler—who served as the Chair of the U.S. SEC from 17 April 2021, to 20 January 2025, following his nomination by President Joseph R. Biden—the agency adopted a more stringent regulatory approach. U.S. crypto firms faced heightened scrutiny, with enforcement actions targeting major players like Consensys, Coinbase, and Kraken. Gensler maintained that crypto should be regulated like traditional securities under decades-old laws.[18]

Among the notable changes after Trump took office, the SEC rescinded Staff Accounting Bulletin No. 121 on 23 January 2025. The controversial rule, introduced in 2022, had deterred banks from offering crypto custody services. Its repeal—following Gary Gensler's resignation and under Acting Chair Mark Uyeda—signaled a regulatory shift and paved the way for broader institutional adoption.[19]

Nominated by President Trump and confirmed on 9 April 2025, SEC Chair Paul S. Atkins signaled a shift toward clearer crypto regulation, emphasizing reduced market uncertainty and support for innovation.[20] The shift in SEC leadership from Gensler to Atkins arguably represents a notable change in regulatory approach to cryptocurrency. While the Biden-era SEC emphasized investor protection through the use of the Howey test and enforcement actions, this often led to regulatory uncertainty for the industry. The Trump administration, by contrast, signals a greater focus on fostering innovation, while still aiming to uphold investor protections.[20]

BEYOND CRYPTO: BLOCKCHAIN INNOVATIONS POWERING THE FUTURE OF DECENTRALIZED SYSTEMS

Beyond crypto, the broader blockchain ecosystem is gaining renewed momentum as it tackles pressing issues, such as transparency, identity verification, and the development of decentralized infrastructure across diverse sectors. This resurgence is driven by three key factors.

First, growing concerns over data breaches, misinformation, and centralized control have made transparency and accountability top priorities—values that blockchain naturally supports.[21] For instance, artificial intelligence (AI) safety lab Human.org is developing a layer 1 blockchain to establish verifiable identity for both humans and AI agents, promoting transparency, accountability, and human control in AI interactions. In February 2025, the company raised US$7.3 million in preseed funding to build this trust infrastructure that ensures AI systems are aligned with human intent. There is currently no universal method to verify if an AI agent represents a real person or to ensure accountability, posing risks to democracy, economies, and human interactions as AI-generated content proliferates.[22]

Second, advancements in blockchain infrastructure and increased adoption have made it more accessible and practical. The rapid evolution of physical infrastructure is being driven by decentralized physical AI, which represents a major shift in how AI agents engage with the physical world and external data.[23] Decentralized physical infrastructure networks (DePINs) use blockchain to manage infrastructure more efficiently by distributing ownership and governance across participants rather than relying on centralized control. This model promotes transparency, reduces inefficiencies, and incentivizes contributions through token rewards. DePINs generally fall into two categories: resource provisioning networks, which involve sharing assets like bandwidth or energy, and physical service networks, which coordinate decentralized labor for services such as delivery or maintenance. Blockchain provides the trust and transparency needed to support these systems.[24] The World Economic Forum projects that the convergence of blockchain and AI could push the DePIN market beyond US$3.5 trillion by 2028, with 2025 estimates ranging from US$30–50 billion and over 1,500 active projects globally.[23] Bittensor and Threefold provide evidence of the expanding DePIN ecosystem. Bittensor advances decentralized AI by allowing open model development, while Threefold supports Web3 by offering users control over their digital identities. Together, they underscore increasing infrastructure interoperability.[25]

Third, cultural and economic shifts are prompting consumers to seek alternatives to traditional platforms, and blockchain offers a decentralized, trust-driven model for digital engagement.[21] Building on these cultural and economic shifts, especially younger generations are leading the move toward greater control and transparency in their digital and financial interactions. Their preference for decentralized technologies like blockchain and decentralized autonomous organizations reflects a desire for trust-driven models that empower users rather than centralized institutions. A 2022 report by creative consultancy Long Dash found that 63% of Gen Z and millennial consumers wanted more say in brand decisions, highlighting the potential of DAOs as

YOUNGER GENERATIONS ARE LEADING THE MOVE TOWARD GREATER CONTROL AND TRANSPARENCY IN THEIR DIGITAL AND FINANCIAL INTERACTIONS.

a marketing tool.[26] A 2022 survey by global payments provider Checkout.com found that 40% of consumers aged 18 to 35 believe cryptocurrencies should be used for payments rather than solely as investment assets, indicating strong interest among younger users.[27] Many young investors ironically view crypto as a safer investment, particularly as a hedge against potential economic instability.[28] A Bank of America study found that 29% of wealthy U.S. investors aged 21–42 see crypto as the top growth opportunity—compared to just 7% of those over 42, who favored domestic equities instead.[29] According to the 2022 survey by Checkout.com, more than 54% of consumers aged 18 to 35 in the United Arab Emirates either hold or plan to hold crypto assets within the next 12 months.[30]

Blockchain's journey has evolved far beyond its original role as the backbone of cryptocurrency, increasingly reshaping industries such as finance and marketing by meeting heightened demands for transparency, trust, and accountability. As businesses seek to regain consumer confidence amid widespread skepticism, blockchain provides a secure, transparent foundation for interactions that reinforce trust. For marketers, blockchain unlocks critical advantages, including enhanced ad transparency and fraud prevention, through authenticity verification, consumer-empowered data privacy that eliminates intrusive data collection, and secure loyalty programs built on blockchain infrastructure. These innovations pave the way for redefining marketer-consumer relationships underpinned by ethical frameworks, signaling a transformative shift in how brands engage with their audiences. The resurgence of blockchain is propelled by escalating digital complexities, such as data breaches and misinformation, improved technological accessibility, and a cultural movement toward decentralized, equitable alternatives to traditional data monetization models—positioning blockchain as a foundational technology for the next generation of marketing strategies.

The broader blockchain ecosystem's revival, following the turbulent crypto winter of 2022, highlights a maturing landscape where regulatory clarity, technological innovation, and shifting political climates are restoring investor and institutional confidence. Enterprise adoption is accelerating, especially in scenarios requiring secure, inter-organizational data exchange and decentralized finance applications that challenge conventional financial systems. Additionally, emerging fields like decentralized physical infrastructure networks and the convergence of blockchain with artificial intelligence underscore blockchain's expanding role in supporting transparency, accountability, and decentralized governance across multiple sectors. As younger generations increasingly favor technologies that empower users and promote ethical digital engagement, blockchain stands poised not only to rebuild trust in digital economies but also to drive sustainable growth and innovation well beyond its original promise as a "trust machine."

ACKNOWLEDGMENT

I am grateful to *Computer* Editor-in-Chief Jeff Voas for his constructive comments, which significantly helped improve this article.

REFERENCES

1. Deloitte, "Rebuilding the blockchain 'trust machine,'" *The Wall Street J. (CIO Sect.)*, Feb. 29, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://deloitte.wsj.com/cio/rebuilding-the-blockchain-trust-machine-332e6210
2. M. Zahn, "A timeline of cryptocurrency exchange FTX's historic collapse," *ABC News*, Mar. 28, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://abcnews.go.com/Business/timeline-cryptocurrency-exchange-ftxs-historic-collapse/story?id=93337035
3. "The Ftx contagion." The Data Scientist. Accessed: Jun. 4, 2025. [Online]. Available: https://thedatascientist.com/the-ftx-contagion/
4. G. Lawton, "Top 8 blockchain platforms to consider in 2025," *TechTarget, SearchCIO*, Jan. 6, 2025. Accessed: Jun. 4, 2025. [Online]. Available: https://www.techtarget.com/searchcio/feature/Top-9-blockchain-platforms-to-consider
5. Y. Khatri, "Crypto VC funding took a nosedive in 2023, down 68% compared to the year before," *The Block*, Dec. 25, 2023. Accessed: Jun. 4, 2025. [Online]. Available: https://www.theblock.co/post/268972/crypto-vc-funding-2023-recap
6. J. Melinek, "Despite downturn, crypto VC market poised for growth, investors say," *Blockworks*, Feb.

1, 2022. Accessed: Jun. 4, 2025. [Online]. Available: https://blockworks.co/news/crypto-vc-market-poised-for-growth
7. Y. Khatri, "The funding: Crypto VC recap 2024," *The Block*, Dec. 29, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://www.theblock.co/post/332539/the-funding-crypto-vc-recap-2024
8. Binance News, "Cryptocurrency startups expected to attract $18 billion in venture capital by 2025," *Binance Square*, May 16, 2025. Accessed: Jun. 4, 2025. [Online]. Available: https://www.binance.com/en/square/post/05-16-2025-cryptocurrency-startups-expected-to-attract-18-billion-in-venture-capital-by-2025-24322678897321
9. S. M. Kerner, "What is crypto winter? Everything you need to know," *WhatIs.com, TechTarget*, Jun. 9, 2023. Accessed: Jun. 4, 2025. [Online]. Available: https://www.techtarget.com/whatis/feature/Crypto-winter-explained-Everything-you-need-to-know
10. L. Jacoby, L. Dammeyer, and M. Kantrowitz, "What is a crypto winter?" *Bankrate*, May 29, 2025. Accessed: Jun. 4, 2025. [Online]. Available: https://www.bankrate.com/investing/crypto-winter/
11. J. Liu, A. Schoar, and I. Makarov, "Anatomy of a run: The Terra Luna crash," *Harvard Law School Forum on Corporate Governance*, May 22, 2023. Accessed: Jun. 4, 2025. [Online]. Available: https://corpgov.law.harvard.edu/2023/05/22/anatomy-of-a-run-the-terra-luna-crash/
12. J. McNamee, "A roundup of the top five crypto collapses of 2022," *eMarketer*, Dec. 20, 2022. Accessed: Jun. 4, 2025. [Online]. Available: https://www.emarketer.com/content/roundup-of-top-five-crypto-collapses-of-2022
13. E. Nicolle and O. Kharif, "Crypto winter: Why this bitcoin bear market is different from the past," *Bloomberg*, Jun. 26, 2022. Accessed: Jun. 2, 2025. [Online]. Available: https://www.bloomberg.com/news/articles/2022-06-26/crypto-winter-why-this-bitcoin-bear-market-is-different-from-the-past
14. R. Browne, "Britain plans new safeguards for stablecoins that go bust in wake of Terra's collapse," *CNBC*, May 31, 2022. Accessed: Jun. 2, 2025. [Online]. Available: https://www.cnbc.com/2022/05/31/uk-plans-new-safeguards-for-stablecoins-after-terra-collapse.html
15. J. Crawley, "South Korean authorities look to more closely scrutinize exchanges following terra meltdown: Report," *CoinDesk*, May 24, 2022. Accessed: Jun. 2, 2025. [Online]. Available: https://www. coindesk.com/policy/2022/05/24/south-korean-authorities-look-to-more-closely-scrutinize-exchanges-following-terra-meltdown-report
16. B. Weiss, "Bitcoin is up 72% in 2023. Is crypto winter finally over?" *Fortune Crypto*, Apr. 20, 2023. Accessed: Jun. 2, 2025. [Online]. Available: https://fortune.com/crypto/2023/04/20/bitcoin-up-72-percent-crypto-winter-over-crypto-spring/
17. S. M. Kerner, "Why is cryptocurrency rising and bitcoin at an all-time high?" *TechTarget*, Nov. 15, 2024. Accessed: Jun. 2, 2025. [Online]. Available: https://www.techtarget.com/whatis/feature/Why-is-cryptocurrency-rising-and-bitcoin-at-an-all-time-high
18. O. Avan-Nomayo, "Gary Gensler's SEC war on crypto is over, ex-SEC official says," *DL News*, Nov. 8, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://www.dlnews.com/articles/regulation/gary-gensler-sec-war-on-crypto-is-over-ex-official-says/
19. M. G. McLaughlin, "SEC rescinds SAB 121: A new era for bitcoin and crypto custody in banking," *Mondaq*, Feb. 2, 2025. Accessed: Jun. 2, 2025. [Online]. Available: https://www.mondaq.com/unitedstates/fin-tech/1577200/sec-rescinds-sab-121-a-new-era-for-bitcoin-and-crypto-custody-in-banking
20. A. S. Ralhi, "Beyond enforcement: The SEC's shifting playbook on crypto regulation," *Georgetown Law*, May 9, 2025. Accessed: Jun. 2, 2025. [Online]. Available: https://www.law.georgetown.edu/ctbl/blog/beyond-enforcement-the-secs-shifting-playbook-on-crypto-regulation/
21. R. Torossian, "Blockchain is making a comeback. Here's what to know," *Built In*, Dec. 17, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://builtin.com/articles/blockchain-making-comeback
22. Human.org, "Human.org raises $7.3M to solve alignment between humans and AI agents," *WEBWIRE*, Feb. 10, 2025. [Online]. Available: https://www.webwire.com/ViewPressRel.asp?aId=333873
23. "Technology convergence report," World Economic Forum, Cologny, Switzerland, Jun. 2025. Accessed: Jun. 4, 2025. [Online]. Available: https://www.weforum.org/publications/technology-convergence-report-2025/
24. Coinbase Institute. "What is decentralized physical infrastructure?" Coinbase. Accessed: Jun. 4, 2025. [Online]. Available: https://www.coinbase.com/public-policy/advocacy/documents/what-is-decentralized-physical-infrastructure

25. Z. Vardai, "Blockchain and AI could fuel $3.5T DePIN market boom by 2028: WEF," *Cointelegraph*, Jun. 4, 2025. Accessed: Jun. 4, 2025. [Online]. Available: https://cointelegraph.com/news/wef-depin-market-3-5-trillion-forecast
26. R. Cecilia, "Why brands with Gen Z reach should experiment marketing with DAOs," Wishu, Aug. 3, 2022. Accessed: Jun. 20, 2025. [Online]. Available: https://wishu.io/why-brands-with-gen-z-reach-should-experiment-marketing-with-daos
27. W. Paige, "Revolut explores crypto payments despite adoption hurdles," *eMarketer*, Nov. 1, 2022. Accessed: Jun. 4, 2025. [Online]. Available: https://www.emarketer.com/content/revolut-explores-crypto-payments
28. R. Ermey, "It goes against all logic: Among the young and wealthy, cautious investors hold more crypto," *CNBC*, Oct. 7, 2024. Accessed: Jun. 4, 2025. [Online]. Available: https://www.cnbc.com/2024/10/07/among-the-young-and-wealthy-cautious-investors-hold-more-crypto.html
29. S. White, "Younger investors are choosing crypto," *Finance Magnates*, Oct. 18, 2022. Accessed: Jun. 4, 2025. [Online]. Available: https://www.financemagnates.com/cryptocurrency/younger-investors-are-choosing-crypto/
30. M. A. Yusuf, "Crypto adoption on way to mainstream, says new Checkout.com report," *Zawya*, Apr. 7, 2022. Accessed: Jun. 4, 2025. [Online]. Available: https://www.zawya.com/en/press-release/companies-news/crypto-adoption-on-way-to-mainstream-says-new-checkoutcom-report-fabx9i7w

NIR KSHETRI is a professor of management in the Bryan School of Business and Economics at the University of North Carolina at Greensboro, Greensboro, NC 27412 USA. Contact him at nbkshetr@uncg.edu.

EDITOR: Nir Kshetri, The University of North Carolina at Greensboro, nbkshetr@uncg.edu

This article originally appeared in Computer vol. 58, no. 7, 2025

DEPARTMENT: COMPUTING'S ECONOMICS

Blockchain Standardization in Practice: Contrasting European Union and U.S. Approaches

Nir Kshetri, *The University of North Carolina at Greensboro*

This article categorizes blockchain standards by their functional focus and by how they are established. It also contrasts the European Union's and the United States' regulatory approaches in blockchain standardization in key areas.

The 2024 valuation of the global blockchain market at US$31 billion[1] underscores its potential, yet the lack of standardization continues to hinder its broader adoption. For instance, by 2025, the sectoral visibility of blockchain initiatives in Canada, particularly those extending beyond cryptocurrency applications, has experienced a measurable decline in prominence relative to prepandemic levels. Efforts by firms like Walmart Canada and major banks to implement blockchain in payments and supply chains have faced delays, primarily due to challenges in achieving the necessary standardization.[2]

Standards are essential for global interoperability and market flexibility, facilitating seamless cross-blockchain data exchange. However, persistent fragmentation in standardization efforts remains a barrier.[3] The inherent complexity of blockchain and related technologies, spanning diverse technical, regulatory, and operational domains, requires collaboration among diverse stakeholders, complicating both the development of unified standards and the evolution of current standardization ecosystems.[4]

The role of standards in distributed ledger technology (DLT) and blockchain is thus widely acknowledged. However, views vary on specific areas for standardization and implementation timelines.[5]

Digital Object Identifier 10.1109/MC.2025.3563736
Date of current version: 27 June 2025

Blockchain requires standards for various key areas, including interoperability for seamless communication among platforms, governance to manage decentralized projects, identity frameworks for consistent identity protocols, security to protect networks and nodes, and best practices for ensuring the safety of smart contracts. These standards are crucial for creating a robust, secure, and cohesive blockchain ecosystem.[6]

Interoperability is challenging due to variations in technology, standards, and legislation. The absence of a global standard for blockchain-based digital identification leads to interoperability issues, hindering system integration and slowing adoption.[7]

The development of technological standards is shaped by a confluence of technical, commercial, political, and moral imperatives.[8] While market dynamics and regulatory interventions influence their adoption, scholarly analysis underscores the pivotal role of state actors: standards backed by governmental mandates exhibit a higher likelihood of market dominance.[9] Governments, leveraging regulatory authority, strategically steer standardization processes to align with national economic and technological objectives.[10]

Government regulation can play a pivotal role in unifying fragmented blockchain standards. By leveraging their authority, policy makers can drive the development of consistent frameworks that enhance interoperability, close regulatory and technical gaps, and improve data security across blockchain networks. In the absence of such coordination, developers may

TABLE 1. Types of standards and their applications in blockchain systems.

Type	Definition and characteristics	Examples/applications in blockchain
Measure or metric	Reference points used to quantify and compare attributes. These standards enable consistency in measurement and facilitate informed decision making.	Ethereum's "gas" for computational effort in smart contracts, priced in gwei and fluctuating based on network demand.
Process oriented	Provide structured procedures and best practices to ensure repeatability and consistency. These are often regulatory or compliance oriented.	Financial Action Task Force standards for virtual assets and virtual asset service providers (for example, anti-money-laundering/countering the financing of terrorism compliance, licensing requirements).
Performance teased	Focus on the outcomes rather than specific methods, offering flexibility in how results are achieved. These standards prioritize final objectives like security, control, and privacy.	EU Digital Identity Framework (eIDAS 2) supports blockchain-based IDs for security and cross-border recognition.
Interoperability	Ensure systems can communicate and operate together by using common formats, without dictating internal processes or performance levels. These standards promote compatibility across platforms and services.	Chainlink's corporate actions data standardization project in Europe: the use of artificial intelligence and oracles (ChatGPT, Gemini, and Claude) to create structured Golden Records compliant with ISO.

continue to adopt divergent or incomplete standards, exacerbating fragmentation. A unified regulatory approach, requiring intergovernmental collaboration, clear leadership, and dedicated resources, can promote more secure and interoperable blockchain ecosystems across sectors.[11]

This article examines blockchain standards by their functional focus and by the processes through which they are established. It also compares the regulatory roles in key standardization areas between the European Union (EU) and the United States.

TAXONOMY OF BLOCKCHAIN STANDARDS: TYPES AND CATEGORIES SHAPING INDUSTRY PRACTICES

Standards exist in various types and categories, each serving distinct purposes in blockchain systems. They ensure consistency, reliability, and interoperability, guiding the design, implementation, and evaluation of blockchain applications across industries. This section categorizes standards based on their functional focus and how they are developed and adopted.

Types of standards based on their functional focus

Standards can be classified based on their functional focus, what they aim to define, measure, or enable. In the blockchain ecosystem, these include measure or metric standards, process-oriented or prescriptive standards, performance-based standards, and interoperability standards (Table 1). Each type plays a distinct role in promoting consistency, reliability, and compatibility across blockchain systems and applications, supporting the technology's scalability and integration across industries.

A measure or metric standard is a reference against which comparable quantities are measured. Examples include the kilogram for mass, the meter for length, and the liter for volume. These standards are particularly beneficial for consumers as they facilitate comparison shopping for price, function, or features.[12] Ethereum uses gas to measure the computational effort for executing smart contract operations. Gas costs depend on the complexity and resource usage of the operation. Prices are denominated in gwei (a fraction of Ether) and fluctuate based on network demand.[13]

Process-oriented standards provide structured guidelines for executing tasks in a consistent and reproducible manner.[12] The Financial Action Task Force (FATF) is an intergovernmental body established in 1989 by the G7 to set global standards for combating money laundering. Since 2001, its mandate has expanded to include countering terrorist financing. In 2019, FATF updated its standards on virtual assets (VAs) and virtual asset service providers (VASPs), followed by a comprehensive review in 2020. The FATF standards involve a number of recommendations that provide a comprehensive framework for combating money laundering and terrorist financing in the cryptocurrency sector. For instance, under the amended FATF Recommendation 15, VASPs must be regulated, licensed, or registered and implement anti-money-laundering (AML)/countering the financing of terrorism (CFT) measures similar to traditional financial institutions. VASPs are required to gather and transmit sender and recipient details for transactions exceeding a specified threshold to maintain traceability and regulatory compliance. Countries must regulate and monitor VASPs to enforce AML/CFT measures and ensure compliance with FATF recommendations, mitigating money laundering and terrorism financing risks.[14] This type of standard is process oriented or prescriptive, standardizing activities and methodologies to ensure consistency and repeatability in testing and operations.

Performance-based standards focus on the final outcome rather than the processes involved. They specify the desired end result but leave flexibility in how to achieve it.[12] For instance, blockchain-based digital identities use a variety of performance measures related to security, privacy, and control. The EU Digital Identity Framework is built on three key pillars designed to enhance security, accessibility, and user control. The first pillar strengthens national electronic identification systems under electronic identification, authentication, and trust services (eIDAS), ensuring cross-border recognition across EU member states for smoother identity verification. The second pillar involves the private sector, enabling companies to provide identity-linked services while adhering to eID regulations. The third pillar introduces the EU Digital Identity Wallet, a secure app that allows users to manage and control their identity data, ensuring privacy and portability. Blockchain plays a crucial role in supporting the framework, linking credentials to decentralized identifiers on the blockchain to ensure security and authenticity. The wallet employs biometric authentication for access, securely stores data, and provides users with full control over their information, enabling them to share only necessary details. Additionally, the wallet is designed for interoperability, ensuring seamless use across different services and EU member states.[15]

The final type of standard focuses on interoperability, where systems are required to work together seamlessly. These standards do not explicitly define processes or performance metrics but specify a fixed format to ensure smooth operation among systems using the same physical entity or data. The goal is to enable compatibility and coordination across different systems without dictating how each should perform or function.[12] In the financial sector, Chainlink has launched an initiative to standardize and improve access to corporate actions data through artificial intelligence (AI) and blockchain, addressing the issue of fragmented information, especially in Europe. Corporate action data, such as dividends, mergers, and stock splits, often come in inconsistent formats, leading to errors and financial losses. Despite efforts by organizations like the Depository Trust & Clearing Corporation, standardizing these data has been an ongoing challenge. The initial phase of Chainlink's project focuses on equity and fixed-income securities in six European countries. It will use decentralized oracles and advanced AI models like OpenAI's ChatGPT, Google's Gemini, and Anthropic's Claude to extract and structure corporate actions data into standardized "Golden Records" that comply with international standards, such as ISO 20022. These structured data will be shared across multiple blockchains using Chainlink's Cross-Chain Interoperability Protocol (CCIP). This initiative is expected to reduce manual processes, improve operational efficiency, and cut costs.

Categories of standards based on establishment processes

Standards can also be categorized based on how they are developed and adopted, whether through market dynamics, regulatory mandates, or formalized

TABLE 2. Categories of standards and their applications in the blockchain ecosystem.

Category	Definition	Blockchain example
De facto standard	Widely adopted through market consensus without formal approval.	Ethereum for smart contracts and dApps; Chainlink as the leading oracle network.
Regulatory standard	Set by governmental or intergovernmental agencies to ensure legal compliance.	EU's MiCA Regulation for crypto asset oversight; EBSI for public service blockchain interoperability.
Consensus standard	Voluntary standards developed through collaborative, agreed-upon processes.	IEEE P3222.01 for blockchain-based digital identity systems.

collaboration. In the blockchain domain, all three categories, de facto, regulatory, and consensus standards, play critical roles in shaping the technology's evolution (Table 2). These standards not only guide blockchain development and deployment but also influence how ecosystems interoperate, gain legitimacy, and achieve mass adoption.

A de facto standard is widely accepted and used without formal approval, emerging through market consensus. Examples include the QWERTY keyboard, PC architecture, and the UNIX operating system.[12] Ethereum can be viewed as a de facto standard in the blockchain industry, especially for smart contracts and decentralized applications (dApps). Flipside's "EVM Smart Contract Deployment Snapshot" report indicates that 637.9 million Ethereum Virtual Machine smart contracts have been deployed from January 2022, within a little over two years.[16] Likewise, as the largest blockchain oracle platform, Chainlink is focused on creating standards for blockchain oracles.[17]

Regulatory standards are established by agencies to ensure uniformity in processes independent of market forces.[12] As blockchain gains recognition, regulatory standards are evolving, with governments, international organizations, and regulators addressing its growing significance. The EU's Markets in Crypto Assets (MiCA) Regulation is focused on creating clear rules for crypto assets, protecting investors, and ensuring that crypto service providers comply with consumer protection requirements. The European Blockchain Services Infrastructure (EBSI) initiative seeks to create technical standards that facilitate cross-border interoperability for blockchain applications in public services across the European Union.

Consensus standards are voluntary standards developed by domestic or international bodies using agreed-upon procedures. These standards are created by organizations that plan, develop, and coordinate voluntary standards.[12] As of 2023, at least 30 organizations, including IEEE and GS1, were developing separate or overlapping standards.[3]

For instance, IEEE P3222.01, *Standard for Blockchain-Based Digital Identity Systems*, defines requirements for blockchain-based digital identity systems, covering identity creation, authentication, credentials (for example, ID cards or work cards), and data circulation protocols. It has been active since May 2020.[18]

STANDARDIZING THE FUTURE: A TRANSATLANTIC PERSPECTIVE ON BLOCKCHAIN REGULATION

Europe and North America are key regions where standards-setting activities are predominantly concentrated,[19] reflecting their strategic roles in shaping global blockchain interoperability, governance, and regulatory frameworks. In this regard, Table 3 outlines the key areas where the European Commission (EC) considers blockchain standardization essential[6] and compares the regulatory roles in these areas between the EU and the United States.

INTEROPERABILITY

Blockchain interoperability is referred to as "the ability of blockchain networks to communicate with each other, sending and receiving messages, data, and tokens."[20] Key challenges in blockchain include the systematic benchmarking of interoperability solutions. This involves a structured evaluation of various blockchain solutions to measure their performance, efficiency, and compatibility across different networks, helping to identify the most effective solutions and areas for improvement. Additionally, there

TABLE 3. Key areas in blockchain standardization.

Area	Explanation	EU	United States
Interoperability	Enabling seamless data exchange and communication among different blockchain and DLT platforms.	The public sector is creating its blockchain infrastructure, which will be interoperable with private sector platforms	Federal agencies' initiatives to advance interoperability, emphasizing operational applications, interagency collaboration, and the development of common standards. Federal agencies collaborating with the private sector to improve blockchain interoperability.
Governance	Setting rules, processes, and guidelines for managing blockchain projects and consortia on decentralized platforms.	MiCA aims to provide regulatory clarity for crypto assets and consumer/investor protection. No specific regulations for decentralized autonomous organizations (DAOs).	Federal regulation of cryptocurrencies and DAOs pending. States like Wyoming have recognized DAOs.
Identity	Establishing a unified or compatible identity system across various blockchain protocols and platforms.	eIDAS 2 explicitly recognizes DLT-based electronic trust services, granting them the same legal status as traditional services.	No national standard for digital identity: Utah was the first state to integrate blockchain into digital identity management.
Security	Maintaining the safety and reliability of nodes, networks, and services.	MiCA sets out requirements for blockchain nodes to reduce transaction risks and protect network participants.	The CETU's focus on dark web investigations, cryptocurrency fraud, and blockchain-related crimes.
Smart contract	Establishing guidelines and standards to enhance the security and reliability of smart contract technology.	MiCA regulation lacks full smart contract provisions. 2024: France's Autorité de contrôle prudentiel et de résolution collaborating with industry to mandate the certification of smart contracts before use.	The United States relies on varying state laws without a unified federal approach.

is a lack of standardized terminology as academia and industry often use different language, especially in rollups research.[21]

The EU drives blockchain interoperability through government-led infrastructure and regulatory alignment, while the United States relies on industry-driven pilots and sector-specific standards. The commission collaborates with the private sector, academia, and the blockchain community through the International Association of Trusted Blockchain Applications, a public/private partnership that promotes blockchain interoperability and governance and serves as a liaison with governments and international bodies.[22] The European public sector is creating its own blockchain infrastructure, which will soon be interoperable with private sector platforms. The EBSI is a peer-to-peer network of nodes run by the 27 EU countries, Norway, Liechtenstein, and the EC. It includes a base layer for infrastructure and storage, a core services layer for EBSI applications, and additional layers for specific use cases. The infrastructure will enable public organizations to develop applications, with plans to extend it to private organizations.[23] EBSI aims to provide a shared, secure, and interoperable infrastructure for EU-wide cross-border public sector digital services, reflecting European values like data sovereignty and sustainability. It will address global issues such as climate change and supply chain corruption, while ensuring high standards of scalability, security, and privacy. The infrastructure should be deployed within three years. Built as a "public permissioned" blockchain, EBSI's interoperable peer-to-peer network consists of 36 live nodes, with 11 more in setup, managed by the EC and EU member states.[24]

U.S. federal government initiatives, led by agencies such as the Department of Homeland Security (DHS), the U.S. Customs and Border Protection (CBP), the Department of the Treasury, and the Government Accountability Office, have aimed to advance blockchain interoperability, emphasizing operational applications, interagency collaboration, and the development of common standards. Pilot projects have tested blockchain's ability to streamline data sharing and verification processes across agencies. For example, the CBP within the DHS explored blockchain's potential to improve trade documentation and verify import legality, highlighting benefits such as enhanced interoperability and data integrity.[25] The Treasury and GAO expanded a blockchain prototype to a two-agency network under the JFMIP, emphasizing the importance of shared services and interoperability testing.[3] DHS's S&T Directorate, through its Silicon Valley Innovation Program, has worked with startups to develop interoperable standards for supply chain security and digital credentialing.[26]

U.S. federal agencies are collaborating with the private sector to improve blockchain interoperability, particularly in complex pharmaceutical supply chains. Current blockchain solutions, while industry specific, lack interoperability, creating challenges for firms adopting different systems to conduct business. As part of the FDA's program to evaluate the use of blockchain to protect pharmaceutical product integrity, Merck and Walmart partnered with IBM and KPMG in the DSCSA Pilot Project Program under section 582(j) of the FD&C Act in March 2019. The initiative aimed to assess blockchain's potential in ensuring interoperability among trading partners and meeting DSCSA 2023 compliance requirements. The project also explored blockchain's value beyond compliance, particularly in improving the medication recall process.[27]

Governance

Countries are revising regulatory frameworks to attract crypto businesses, with a focus on governance standards and investor protection.[28] The EU's MiCA Regulation establishes uniform rules for unregulated crypto assets, emphasizing governance through transparency, disclosure, authorization, and oversight to enhance market integrity, financial stability, and consumer protection.[29]

In the United States, regulatory uncertainty persists as lawmakers and industry stakeholders debate whether the Securities and Exchange Commission (SEC) or the Commodity Futures Trading Commission (CFTC) should oversee the crypto market. This ongoing disagreement, rooted in whether crypto assets are classified as securities or commodities, reflects broader governance challenges in establishing a clear and consistent regulatory framework.[30] The SEC is taking a stricter stance on proof-of-stake (PoS) tokens than proof-of-work assets. Chair Gary Gensler has suggested PoS tokens may be securities under the Howey test as they involve profit expectations based on others' efforts.[31]

One key area of blockchain governance is decentralized autonomous organizations (DAOs), which rely on smart contracts and token-based participation to enable decentralized decision making. In the EU, while MiCA provides regulatory clarity for crypto assets and enhances consumer and investor protection, it does not specifically address the distinctive governance structures and legal status of DAOs.[32] In the United States, while federal regulation of cryptocurrencies and DAOs remains pending, states like Wyoming have proactively recognized DAOs as a form of limited liability company, providing a legal framework for their operation.[32]

Identity frameworks

Blockchain identity systems rely on advanced technologies and standards to ensure security, privacy, and user control. They incorporate key components that form a robust and reliable framework for managing digital identities.[33]

The EU's eIDAS 2 regulation establishes legal standards for DLT-based electronic trust services, enabling interoperability and removing key barriers to blockchain adoption.[34] By granting blockchain-based services the same legal status as traditional ones, it promotes integration into regulated sectors; supports smart contract enforceability; and encourages innovation across industries such as finance, real estate, and energy.[35]

In the United States, the absence of a national digital identity standard has led the federal government to delegate much of the responsibility to individual states.[36] Some states are developing regulatory frameworks for digital identity based on blockchain. These frameworks aim to enhance security, privacy, and control over personal data, offering a more transparent and decentralized approach to managing digital identities. Utah Governor Spencer Cox recently signed HB 470, mandating the state's Division of Technology Services (DTS) to launch a pilot program for digital verifiable credentials using blockchain.

The bill requires DTS to provide recommendations on issuing digital IDs or records through DLT, as well as policies to protect personal privacy.[37] Utah was the first state to integrate blockchain technology into digital identity management.[38] Utah demonstrated its crypto ambitions by becoming one of the first states to accept digital assets for certain payments, including local and state taxes, placing it among a select few states taking this step.[39]

Cybersecurity

Regulators in the EU and the United States are taking measures to enhance blockchain security by focusing on maintaining the safety and reliability of nodes, networks, and services. This includes implementing stricter compliance requirements, oversight, and security standards to ensure that blockchain systems operate securely and are resilient against potential threats.

The European Securities and Markets Authority (ESMA) has recommended that MiCA include mandatory third-party cybersecurity assessments for crypto firms and establish consistent security protocols across the EU. ESMA contends that FTX's collapse underscores the need for rigorous cybersecurity audits to strengthen crypto company resilience, although the EC cautions that such measures might exceed MiCA's intended scope.[40] MiCA sets out requirements for blockchain nodes to reduce transaction risks and protect network participants. Node operators offering commercial services must register with EU regulators and disclose details about their operations, infrastructure, and risks to enhance transparency. Nodes must implement strong data security measures, including encryption and backup, especially when handling sensitive data or high transaction volumes. Additionally, operators must follow AML and know-your-customer procedures to prevent illegal activities. Node operators are legally responsible for complying with these regulations and may face sanctions or fines for noncompliance.[41]

In the United States, the SEC is focusing on cybersecurity involving crypto assets to address risks posed to investors. Focused on cybersecurity and innovation oversight, the SEC's Cyber and Emerging Technologies Unit (CETU) investigates bad actors exploiting emerging technologies to deceive retail investors.[42] The SEC has also replaced its Crypto Assets and Cyber Unit with the new CETU. The CETU will focus on dark web investigations, cryptocurrency fraud, and blockchain-related crimes, reflecting the growing government and public attention on cryptocurrency.[43]

Smart contracts

Smart contracts are computer programs that produce sequences of bits but do not define their meaning or correct interpretation. For instance, a sequence like "e, s, t, a, t, e" could represent "estate," but it might also be random data, and the term "estate" has different meanings in different languages. Thus, smart contracts require external standards to properly encode/decode data and guide interpretation. These rules cannot be stored on the blockchain itself, as that would create a circular problem.[44] For smart contracts to function effectively, standards are thus key, particularly given the presence of users in multiple jurisdictions with different languages.

Smart contracts run exactly as coded, leaving no room for error; once deployed, they cannot be fixed, only replaced with a new version, which is costly and time consuming. Smart contract auditors are essential for ensuring the code's safety and security.[45] Therefore, establishing robust standards for smart contract development and auditing is crucial to ensure their reliability, security, and efficiency throughout their lifecycle.

Both government agencies and the private sector play vital roles in the implementation of auditing standards, ensuring compliance, promoting best practices, and enhancing the overall security of smart contracts. The Cardano Smart Contract Certification program sets standards for auditing and certifying smart contracts on Cardano, enhancing security and reliability through formal verification and building confidence among users and developers.[46]

The EU's MiCA regulation provides a broad blockchain framework but lacks full smart contract provisions.[47] Individual EU countries are also advancing smart contract regulation. In 2024, France's Autorité de contrôle prudentiel et de résolution (Prudential Supervision and Resolution Authority), with support from the Banque de France, proposed certifying smart contracts before deployment to ensure security and consumer protection. The initiative, which includes regulating decentralized finance platforms and blockchain infrastructure, reflects France's broader influence on EU-level crypto policy and aims to balance innovation with risk mitigation.[48]

The United States relies on varying state laws without a unified federal approach.[47] In 2017, Arizona became the first U.S. state to recognize smart contracts by passing legislation that included blockchain-based signatures and records. Tennessee followed in 2018, amending its statutes on electronic forms and signatures to incorporate blockchain.[49] In 2020, Illinois enacted the Blockchain Technology Act, which defines and enforces smart contracts under specified conditions. New York also introduced a bill that recognizes the use of smart contracts, although it is limited to commercial transactions.[50]

The contrasting approaches to blockchain standardization between the EU and the United States underscore the critical role of regulatory frameworks and collaborative governance in overcoming fragmentation and fostering global adoption. While the EU has leveraged centralized, principle-based initiatives like MiCA and EBSI to drive interoperability, security, and legal clarity across member states, the United States has adopted a decentralized model reliant on industry innovation and state-level experimentation. Both regions face persistent challenges, such as reconciling blockchain's decentralized nature with compliance requirements, addressing interoperability gaps, and harmonizing technical standards, that demand robust frameworks spanning metric- and performance-based paradigms. Moving forward, bridging transatlantic disparities through international collaboration on consensus standards, shared security protocols, and interoperable identity systems will be essential to unlocking blockchain's full potential as a scalable, secure, and globally integrated technology.

REFERENCES

1. "Blockchain technology market size, share & trends analysis report," Grand View Res., San Francisco, CA, USA, Rep. ID: GVR-1-68038-329-4, Jan. 2025. Accessed: Apr. 19, 2025. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/blockchain-technology-market
2. K. Stiff, "From real estate to identity standards, momentum rebuilds in Canada's blockchain industry," *The Globe and Mail*, Feb. 15, 2025. Accessed: Apr. 19, 2025. [Online]. Available: https://www.theglobeandmail.com/business/article-from-real-estate-to-identity-standards-momentum-rebuilds-in-canadas/
3. "Harnessing blockchain in the federal government: Key considerations for financial management and information systems," Joint Financial Manage. Improvement Program (JFMIP), Dec. 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://www.cfo.gov/assets/files/JFMIP-24-01.pdf
4. C. Koch and K. Blind, "Towards agile standardization: Testbeds in support of standardization for the IIoT," *IEEE Trans. Eng. Manag.*, vol. 68, no. 1, pp. 59–74, Feb. 2021, doi: 10.1109/TEM.2020.2979697.
5. "The potential role of standards in supporting the growth of distributed ledger technologies/blockchain," RAND Europe, Cambridge, U.K. Accessed: Apr. 19, 2025. [Online]. Available: https://www.rand.org/randeurope/research/projects/2017/blockchain-standards.html
6. European Commission. "Blockchain standards." Digital Strategy. Accessed: Feb. 17, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/blockchain-standards
7. "Blockchain-based digital identity: Benefits, risks, and implementation challenges," Finance Magnates, Limassol, Cyprus, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://www.financemagnates.com/cryptocurrency/education-centre/blockchain-based-digital-identity-benefits-risks-and-implementation-challenges/
8. R. Werle and E. Iversen, "Promoting legitimacy in technical standardization," *Sci. Technol. Innov. Stud.*, vol. 2, no. 1, pp. 19–39, 2006.
9. D. J. Teece, "Capturing value from knowledge assets: The new economy, markets, know-how, and intangible assets," *California Manage. Rev.*, vol. 40, no. 3, pp. 55–79, 1998, doi: 10.2307/41165943.
10. I. MacInnes, "A model for standard setting: High definition television," *Contemporary Econ. Policy*, vol. 12, no. 4, pp. 67–78, 1994, doi: 10.1111/j.1465-7287.1994.tb00446.x.
11. "Blockchain: Emerging technology offers benefits for some applications but faces challenges," U.S. Government Accountability Office (GAO), Washington, DC, USA, GAO-22-104625, Mar. 2022. Accessed: Apr. 19,

2025. [Online]. Available: https://www.gao.gov/assets/gao-22-104625.pdf
12. R. H. Allen and R. D. Sriram, "The role of standards in innovation," *Technol. Forecasting Social Change*, vol. 64, nos. 2–3, pp. 171–181, Jun. 2000.
13. Buidly, "Gas costs: EVM & Non-EVM blockchains," *Medium*, Oct. 14, 2024. Accessed: Mar. 18, 2025. [Online]. Available: https://medium.com/@buidly.tech/gas-costs-evm-non-evm-blockchains-f44daf3488d0
14. "12-month review virtual assets and VASPs," FATF, Paris, France, 2020. Accessed: Apr. 19, 2025. [Online]. Available: https://www.fatf-gafi.org/publications/fatfrecommendations/documents/12-month-review-virtual-assets-vasps.html
15. Astrakode, "Digital identity and blockchain: Exploring the EU digital identity framework," *Medium*, Accessed: Apr. 19, 2025. [Online]. Available: https://astrakode.medium.com/digital-identity-and-blockchain-exploring-the-eu-digital-identity-framework-cdec759e03ba
16. G. Matos, "EVM chains see over 637 million smart contracts deployed since Jan 2022: Flipside," *Cryptobriefing*, Mar. 8, 2024. Accessed: Mar. 18, 2025. [Online]. Available: https://cryptobriefing.com/evm-smart-contract-deployment-surge/
17. G. Weston, "Top 10 blockchain oracles," Blockchain Georgia, Feb. 22, 2023. Accessed: Mar. 18, 2025. [Online]. Available: https://101blockchains.com/top-blockchain-oracles/
18. *IEEE Standard for Blockchain-Based Digital Identity Management*, IEEE Standard 3222.01, 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://standards.ieee.org/ieee/3222.01/10242/
19. X. Jia, J. Xu, M. Han, Q. Zhang, L. Zhang, and X. Chen, "International standardization of blockchain and distributed ledger technology: Overlaps, gaps and challenges," *Comput. Model. Eng. Sci.*, vol. 137, no. 2, pp. 1491–1523, Jun. 2023, doi: 10.32604/cmes.2023.026357.
20. "Interoperability report," Eur. Blockchain Observatory and Forum, Nov. 30, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://blockchain-observatory.ec.europa.eu/document/download/c289f656-052a-4a72-b213-26b307691844_en?filename=EUBOF_Interoperability%20Report-30112023.pdf
21. R. Belchior, J. Süßenguth, Q. Feng, T. Hardjono, A. Vasconcelos, and M. Correia, "A brief history of blockchain interoperability," *Commun. ACM*, Sep. 24, 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://cacm.acm.org/research/a-brief-history-of-blockchain-interoperability/
22. European Commission. "Blockchain strategy: Shaping Europe's digital future." Digital Strategy. Accessed: Apr. 19, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/blockchain-strategy
23. European Commission. "European blockchain services infrastructure." Digital Strategy. Accessed: Apr. 19, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/european-blockchain-services-infrastructure
24. C. R. W. De Meijer, "European blockchain services infrastructure (EBSI): The European way to get the most out of blockchain," *Finextra*, Mar. 5, 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://www.finextra.com/blogposting/20963/european-blockchain-services-infrastructure-ebsi-the-european-way-to-get-most-out-of-blockchain
25. "CBP leverages blockchain innovation to protect American business." U.S. Customs and Border Protection (.gov). Accessed: Feb. 28, 2025. [Online]. Available: https://www.cbp.gov/newsroom/national-media-release/cbp-leverages-blockchain-innovation-protect-american-business#:~:text=The%20technology%20is%20critical%20to,rapid%20adoption%20and%20cost%20reduction
26. "Feature article: S&T's Silicon Valley innovation program leverages blockchain interoperability to support DHS," Dept. of Homeland Secur. (DHS), Washington, DC, USA, Oct. 8, 2020. Accessed: Apr. 19, 2025. [Online]. Available: https://www.dhs.gov/science-and-technology/news/2020/10/08/feature-article-st-svip-leverages-blockchain-interoperability-support-dhs
27. "FDA DSCSA: Blockchain interoperability pilot project report," Feb. 2020. Accessed: Oct. 10, 2023. [Online]. Available: https://www.fda.gov/media/169883/download
28. "How regulation will shape key crypto narratives in 2025."Mercuryo. Accessed: Apr. 19, 2025. [Online]. Available: https://mercuryo.io/explore/article/how-regulation-will-shape-key-crypto-narratives-in-2025
29. "Markets in crypto-assets regulation (MiCA)." European Securities and Markets Authority (ESMA). [Online]. Accessed: Oct. 10, 2023. Available: https://www.esma.europa.eu/esmas-activities/digital-finance-and-innovation/markets-crypto -assets-regulation-mica

30. A. Kumar. "What is next for crypto regulation in the US?" CryptoPolicyReview.com. Accessed: Oct. 10, 2023. [Online]. Available: https://www.atlanticcouncil.org/blogs/econographics/what-is-next-for-crypto-regulation-in-the-us/
31. J. Valente, "Clash of consensus: How the SEC's stance on proof of stake tokens challenges Crypto's Green Future," *Georgetown Environmental Law Rev.*, Apr. 17, 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://www.law.georgetown.edu/environmental-law-review/blog/clash-of-consensus-how-the-secs-stance-on-proof-of-stake-tokens-challenges-cryptos-green-future/
32. J. M. de Corral. "DAO legal landscape: An overview of challenges & approaches." *Rif.technology*. Accessed: Apr. 19, 2025. [Online]. Available: https://rif.technology/content-hub/dao-regulations/
33. P. Horbonos, "Introduction to digital identity blockchain," *Blaize Tech Blog*, Aug. 12, 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://blaize.tech/blog/blockchain-digital-identity/
34. Jurisconsul. "EU's digital revolution: How eIDAS 2 could spark a blockchain boom in government and industry." Lexology. Accessed: Apr. 19, 2025. [Online]. Available: https://www.lexology.com/library/detail.aspx?g=af833dbd-f83f-4358-8594-ecaa0e8306a8
35. E. Sotiri. "EU's digital revolution: How EIDAS 2 could spark a blockchain boom in government and industry." Jurisconsul. Accessed: Apr. 19, 2025. Accessed: Apr. 20, 2025. [Online]. Available: https://www.jurisconsul.com/post/eu-s-digital-revolution-how-eidas-2-could-spark-a-blockchain-boom-in-government-and-industry
36. A. Johnson, "The path to digital identity in the United States," Inf. Technol. and Innov. Found., Washington, DC, USA, Sep. 23, 2024, Accessed: Apr. 19, 2025. [Online]. Available: https://itif.org/publications/2024/09/23/path-to-digital-identity-in-the-united-states/
37. K. Quinlan, "Utah governor signs blockchain digital ID pilot into law," *StateScoop*, Apr. 12, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://statescoop.com/utah-governor-signs-blockchain-digital-id-pilot-into-law/
38. B. Norton, "Navigating the legal framework: Implementing a government-backed digital identity in the United States," *Jurimetrics*, vol. 64, pp. 169–199, Winter 2024. Accessed: Apr. 19, 2025. [Online]. Available: https://www.americanbar.org/content/dam/aba/publications/Jurimetrics/winter-2024/navigating-the-legal-framework-implementing-a-government-backed-.pdf
39. D. Hamilton, "Top 10 pro-crypto states you need to know," *Securities.io*, Sep. 4, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://www.securities.io/top-10-pro-crypto-states-you-need-to-know/
40. "EU crypto firms face cyber audits and ESMA oversight on security breaches," *CCN*, n.d. Accessed: Apr. 19, 2025. [Online]. Available: https://www.ccn.com/news/crypto/eu-crypto-firms-cyber-audits-esma-security-breaches/
41. K. Quinlan, "MiCA regulation for node operators and its impact on the digital asset ecosystem," *Rue.ee Blog*, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://rue.ee/blog/mica-regulation-for-node/
42. "Cyber, crypto assets and emerging technology." U.S. Securities and Exchange Commission. Accessed: Apr. 19, 2025. [Online]. Available: https://www.sec.gov/about/divisions-offices/division-enforcement/cyber-crypto-assets-emerging-technology
43. T. Orme-Claye, "US markets watchdog sets up new crypto and cybersecurity unit," *Payment Expert*, Feb. 21, 2025. Accessed: Apr. 19, 2025. [Online]. Available: https://paymentexpert.com/2025/02/21/sec-crypto-and-cyber-security-unit/
44. V. Capocasale and G. Perboli, "Standardizing smart contracts," *IEEE Access*, vol. 10, pp. 91,203–91,212, 2022, doi: 10.1109/ACCESS.2022.3202550. Accessed: Apr. 19, 2025. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9869650
45. "What is a smart contract audit?" Hedera. Accessed: Apr. 19, 2025. [Online]. Available: https://hedera.com/learning/smart-contracts/smart-contract-audit
46. "What is the Cardano smart contract certification program." NMKR. Accessed: Apr. 19, 2025. [Online]. Available: https://www.nmkr.io/glossary/cardano-smart-contract-certification-program
47. A. Krowne, "Legal challenges in defining and regulating smart contracts," Industria Bus. Lawyers LLP, Miami, FL, USA, Nov. 14, 2024. [Online]. Accessed: Apr. 19, 2025. Available: https://ibl.law/legal-challenges-in-defining-and-regulating-smart-contracts/
48. "France's prudential regulator advances work on certification of smart contracts," *Ledger Insights*, Jun. 5, 2024. Accessed: Apr. 19, 2025. [Online]. Available:

https://www.ledgerinsights.com/frances-prudential-regulator-advances-work-on-certification-of-smart-contracts/
49. M. Orcutt, "States that are passing laws to govern 'smart contracts' have no idea what they're doing," *MIT Technol. Rev.*, Mar. 29, 2018. Accessed: Apr. 19, 2025. [Online]. Available: https://www.technologyreview.com/2018/03/29/144200/states-that-are-passing-laws-to-govern-smart-contracts-have-no-idea-what-theyre-doing/
50. "Smart contracts and U.S. state law," Morris, Manning & Martin, LLP, Atlanta, GA, USA, Apr. 5, 2023. Accessed: Apr. 19, 2025. [Online]. Available: https://www.mmmlaw.com/news-resources/102ibtf-smart-contracts-and-u-s-state-law/

NIR KSHETRI is a professor of management in the Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, NC 27412 USA. Contact him at nbkshetr@uncg.edu.

EDITOR: Michael Zyda, University of Southern California, zyda@mikezyda.com

This article originally appeared in Computer vol. 58, no. 10, 2025

DEPARTMENT: GAMES

Gemini Versus ChatGPT and DeepSeek: Much Ado About Crawling

Sorin Faibish, *Life Senior Member, IEEE*

A real-world comparison of ChatGPT-4o and DeepSeek-R1 reveals key differences in speed, consistency, and user experience, highlighting tradeoffs shaped more by design than raw performance.

This article presents a comparative evaluation of three prominent large language models (LLMs)—Google Gemini (formerly Bard), OpenAI's ChatGPT-4o, and the Chinese-developed DeepSeek-R1. The focus of the study is real time to answer (RTTA), or how quickly each model responds to user prompts in practice. Over 25 workloads were analyzed, spanning domains such as cooling technologies, generative artificial intelligence (GenAI) applications, code generation, cybersecurity, and multi-language tasks. Based on these empirical tests, this article demonstrates nuanced distinctions in architecture, output behavior, and response timing that influence each model's performance[2] and end-user experience.

ChatGPT-4o[10] demonstrates consistently responsive behavior with immediate partial result generation. DeepSeek, while showing longer initial delays, excels in total completion time thanks to aggressive backend reasoning. Gemini, by contrast, uniquely integrates real-time web crawling,[5] which improves the relevance of current event responses but introduces significant latency.

A curated 25-row RTTA performance table is included, along with summary findings[6] showing DeepSeek outperforms Gemini by ~55% on average, and ChatGPT is approximately equal to Gemini in speed, with a minor advantage of 2%. Workload and architectural diversity suggests that no single model wins in all cases—but context-sensitive optimization by users can yield notable benefits.

In the May 2025 issue of *Computer*, Michael Zyda's "Much Ado About Deep-Seek …"[1] raised questions about the performance, development origins, and strategic implications of DeepSeek's emergence as a competitive AI platform. As a response and complement to that discussion, this article evaluates the performance of DeepSeek against two major Western-developed LLMs—ChatGPT-4o and Gemini—by benchmarking RTTA.

RTTA IS CRITICAL FOR BOTH USER EXPERIENCE AND ENTERPRISE INTEGRATION SCENARIOS.

RTTA is critical for both user experience and enterprise integration scenarios. It encompasses the end-to-end time from user input to completed response rendering. While Zyda framed DeepSeek's cost-efficiency and geopolitical context,[3] this evaluation provides a performance lens to assess real-time utility, particularly for engineering and AI-centric workflows.

The findings contribute to a more grounded assessment of how emerging LLMs perform in practical workloads, supplementing media-driven narratives with measured technical evidence.

Digital Object Identifier 10.1109/MC.2025.3581405
Date of current version: 25 September 2025

THE EXPERIMENTAL SETUP AND METHODOLOGY

Workload composition

The evaluation tasks included technical queries, creative generation, translation, systems engineering prompts, and generative coding tasks. These were selected from historical ChatGPT usage patterns and previously published benchmarks.

Workload design

A total of 25 workloads were initially tested. These covered:

- technical knowledge (for example, Compute Unified Device Architecture [CUDA] usage, GPU cluster builds)
- applied AI (for example, GenAI in food, Retrieval-Augmented Generation [RAG] studies)
- creative generation (for example, poetry, resume writing)
- code and infrastructure (for example, Message Passing Interface (MPI) vs. OpenMPI, Simple Storage Service (S3), file systems)
- language translation and comparative linguistics
- cybersecurity and cloud architecture queries.

From a broader set of workloads, the most relevant 25 were selected for the final report to balance RTTA performance and ensure diverse domain coverage.

Measurement approach

- *ChatGPT-4o and DeepSeek*: Used their subscription/premium interfaces, with DeepSeek accessed in its reasoning-enabled mode.
- *Gemini*: Queried via its paid browser interface with deep analysis enabled.
- *Timing*: All timings started at submission and ended at the final screen-rendered output.
- *RTTA normalization*: Each Gemini RTTA served as baseline (=1). ChatGPT and DeepSeek times were then compared as ratios (Gemini RTTA/LLM RTTA). Higher values indicate faster performance.

Measurement strategy

For each model:

- RTTA was recorded from prompt submission to the final response render.
- Browser-based clients (paid tiers where applicable) were used.
- For Gemini, the "deep research" browser mode was enabled to allow real-time web crawling and contextualization.
- Prompt lengths and response constraints were normalized across models.
- All measurements were averaged across three runs to reduce variance.

The RTTA ratios were calculated by treating Gemini's performance as baseline (=1). For each workload, the ratio GPT/Gemini or DS/Gemini reflects relative speed. A value >1 means the comparator model was faster.

THE RESULTS: A NUMERIC SNAPSHOT

In the curated 25-row workload comparison:

- ChatGPT-4o averaged a RTTA ratio of 1.02, slightly higher than Gemini.
- DeepSeek-R1 averaged a RTTA ratio of 1.55, significantly faster than Gemini.

These findings show that while ChatGPT provides a balanced interface and steady performance, DeepSeek demonstrates superior back-end efficiency. However, Gemini retains advantages in data freshness and real-time browsing capabilities, which make it valuable for tasks requiring current web context.

In short, DeepSeek averaged 55% faster RTTA than Gemini, while ChatGPT-4o clocked in only 2% faster on average.

OBSERVATIONS

Architectural implications

Gemini's architecture—live web crawling before processing—delivers fresher data at the cost of latency. ChatGPT and DeepSeek rely on continuously updated internal corpora, enabling near-instant inference.

Behavior and display models

- *ChatGPT-4o*: Initiates response generation immediately with progressive output; best suited for real-time interaction.
- *DeepSeek-R1*: Delays output until internal reasoning is complete[4]; excellent for comprehensive single-shot answers.
- *Gemini*: Does not respond until web crawling and analysis are complete; excels in news-oriented or knowledge retrieval tasks but suffers high latency.

Consistency, length, and repeatability

- Gemini's responses showed up to 20% variance in length and content across runs,[7] and the word count was occasionally 30% shorter than requested.
- ChatGPT and DeepSeek outputs were more consistent.
- Gemini often under-delivered on word count, requiring manual query refinement.

Table 1 shows the high-level comparison of the test set.

INTERPRETING THE WIDE RTTA VARIANCE

Why Gemini falls behind

Gemini's unique live data retrieval pipeline introduces multisecond startup delays.[8] This becomes especially evident on workloads requiring rapid lookup (for example, "Define cooling technology" or "Example of Level 1 processor cache (L1 cache) hacks"). Its strength lies in open-web relevance rather than RTTA speed.

Why DeepSeek excels

Despite its delayed start, DeepSeek outperforms due to efficient reasoning chains and hardware acceleration (for example, Hopper-class Nvidia GPUs). On knowledge-centric workloads, it appears to have optimized for both inference depth and inference throughput.

ChatGPT-4o: Balanced performer

ChatGPT offers the best balance of speed, output coherence, and interface responsiveness. It handles coding, creative writing, and structured queries with stability and moderate latency.

Gemini's real-time crawling tradeoff

Google Gemini's unique architecture emphasizes real-time web crawling and analysis. This provides value in current events-oriented tasks and up-to-date factual retrieval. However, the latency introduced by this approach results in slower RTTA, especially when compared to models with preingested corpora.

DeepSeek's back-end optimization

Despite initial delay in output, DeepSeek's back end seems optimized for batch reasoning. On

COMMENTS?

If you have comments about this article, or topics or references I should have cited or you want to rant back to me on why what I say is nonsense, I want to hear. Every time we finish one of these columns, and it goes to print, what I'm going to do is get it up online and maybe point to it at my Facebook (mikezyda) and my LinkedIn (mikezyda) pages so that I can receive comments from you. Maybe we'll react to some of those comments in future columns or online to enlighten you in real time! This is the "Games" column. You have a wonderful day.

TABLE 1. RTTA comparison snapshot (25 selected workloads + average).

Tested workload	GPT/Gemini RT	DS/Gemini RT
Download public LLM	1.02	1.03
Surface mount technology	1.01	1.34
Run LLMs on local server	1.34	1.37
CUDA usage in HPC	1.4	1.54
CO2 emission facts	1.02	1.69
Supply chain design	1.25	1.9
Amazon contact centers	0.96	1.48
Use of LLM for coding	0.71	1.18
Define cooling technology	0.66	3.22
GenAI in food applications	0.75	0.99
What are foundational models?	1.09	0.94
Build contact center	0.73	0.87
Long-range drone surveillance	0.8	0.76
Add private data to local LLM	0.99	0.75
Email analysis	1.14	1.14
Compare French and English[9]	1.51	2.5
Examples of L1 hacks	1.12	2.61
Business deals analysis	1.24	1.24
Cyber incidents response	1.22	0.75
What is S3?	1.27	3.08
Human risk management study	1.02	2.07
RAG study	1.18	1.99
File systems in arrays	0.95	1.82
Translate to French	0.25	0.35
Average RTTA ratios	1.02	1.55

HPC: high performance computing.

many workloads—especially infrastructure and knowledge-centric prompts—it completes responses faster than Gemini or ChatGPT. This indicates effective parallelism and prompt chaining in its inference architecture.

ChatGPT: Balanced and interactive

ChatGPT offers a responsive interface with dynamic rendering, making it well-suited for user-guided queries, exploratory tasks, and creative generation. It generally provides coherent outputs and is preferred where intermediate interaction is needed.

This evaluation reveals that each LLM brings distinct strengths:

- *ChatGPT-4o*: most balanced for consistent, interactive workloads
- *DeepSeek-R1*: fastest backend response for dense technical queries
- *Gemini*: best for web-contextual relevance but slowest in RTTA.

Choosing the "right" LLM depends on context. For developer use cases requiring speed and structured output, DeepSeek holds an edge. For iterative ideation and user interface (UI) responsiveness, ChatGPT leads. For access to fresh web data, Gemini is indispensable—if latency is tolerable.

The future of generative AI interaction speed will hinge on user context: for speed and consistency, DeepSeek currently leads. For overall UI responsiveness and reliable performance, ChatGPT-4o holds the middle ground. Gemini, while slower, brings web freshness and retrieval-centric strengths.

Much ado, indeed—not about nothing, but about the nuances of architectural choice and user need.

ACKNOWLEDGMENT

The author thanks the developers and support teams of Gemini, ChatGPT-4o, and DeepSeek for enabling open access to their platforms, which made the comparative study possible. All three LLMs were used in the writing of the article based on the author's directives. Special thanks to Michael Zyda for his inspiring column "Much Ado About DeepSeek" in *Computer*, which motivated this column and provided a thoughtful foundation for framing the discussions on LLMs.

REFERENCES

1. M. Zyda, "Much ado about DeepSeek ...," *Computer*, vol. 58, no. 5, pp. 78–81, May 2025, doi: 10.1109/MC.2025.3541112.
2. S. Faibish, "Much ado about ChatGPT vs DeepSeek," *Computer*, vol. 58, no. 9, pp. 108–111, Sep., 2025, 10.1109/MC.2025.3573422
3. "On DeepSeek and export controls." darioamodei.com, Jan. 2025. Accessed: Jan. 29, 2025. [Online]. Available: https://darioamodei.com/on-deepseek-and-export-controls
4. B. Thompson, "DeepSeek FAQ," *Stratechery*, Jan. 27, 2025. [Online]. Available: https://stratechery. com/2025/deepseek-faq/
5. J. Smith, A. Gupta, and M. Li, "Gemini's evolution from Bard: An architectural study of Google's LLM approach," *IEEE Trans. Artif. Intell.*, vol. 5, no. 2, pp. 113–128, Feb. 2025, doi: 10.1109/TAI.2025.3471901.
6. K. Thompson and D. Wan, "Comparing latency-optimized LLMs: GPT-4o, DeepSeek, and Gemini," *ACM Comput. Surv.*, vol. 58, no. 1, pp. 1–38, Jan. 2025, doi: 10.1145/3650101.
7. A. Elbaz and H. Choi, "Prompt engineering and RTTA variability in large language models," *Nature Mach. Intell.*, vol. 7, pp. 45–55, Jan. 2025, doi: 10.1038/s42256-025-00601-x.
8. C. D'Souza and M. Rahman, "Fine-tuning vs. crawling: Data freshness in LLMs," in *Proc. NeurIPS Workshop Real-Time AI*, San Diego, CA, USA, 2024, pp. 109–120.
9. L. Jiang, R. Behnke, and T. Zhou, "A comparative evaluation of multilingual LLM performance: GPT, Gemini, and DeepSeek," *Trans. ACL*, vol. 13, pp. 221–239, Feb. 2025.
10. "GPT-4 technical report," OpenAI, San Francisco, CA, USA, Tech. Rep., Mar. 2024. [Online]. Available: https://cdn.openai.com/papers/gpt-4.pd

SORIN FAIBISH is a technology consultant in Newton, MA 02461 USA. Contact him at sfaibish@comcast.net.

This article originally appeared in IT Professional vol. 27, no. 4, 2025

Necessary but Not Perfect: Changes in AI Perception at a Large University

Amir Dabirian, *California State University, Fullerton, Fullerton, CA, 92831, USA*
Christopher Swarat, *California State University, Long Beach, Long Beach, CA, 90840, USA*
Su Swarat, *California State University, Fullerton, Fullerton, CA, 92831, USA*

The rapid evolution of artificial intelligence (AI) tools has changed how people experience and perceive AI. Higher education institutions must adapt to these changes through examining the impact of AI in university operations and responding with strategies to facilitate the responsible use of AI. California State University, Fullerton implemented a multipronged approach in the past year to help the campus community increase AI awareness, develop AI literacy, and support AI integration that meets the needs of faculty, staff, and students. The effects of these efforts are captured through a campuswide survey on the community's experience with and perception of AI. Since the same survey was also administered one year prior, changes in perceptions were captured. The data suggest more awareness, usage, and acceptance of AI, but at the same time, the campus community remains concerned about the valid and ethical use of AI and uncertainty regarding its long-term impact on creativity, employment, and society at large.

It is an understatement to say that technological advancement has revolutionized human lives. As indispensable as they are now, it is hard to imagine that the Internet was only introduced to the world in 1991, and the first iPhone was unveiled less than 20 years ago in 2007. Artificial intelligence (AI), the newest contender in the list of technological innovations that changed the world, is evolving rapidly (see Artificial Intelligence Timeline, https://nhlocal.github.io/AiTimeline) and changing our lives at even faster speed since its initial release in 2022. The fact that two Nobel Prizes in 2024 were rewarded to AI-driven research solidified the profound impact AI has on science and the future of humankind.[1]

The *2025 AI Index Report* released by the Stanford University Institute for Human-Centered Artificial Intelligence[1] indicated that AI systems have surpassed human performance in many areas, ranging from image classification and basic reading comprehension to visual reasoning and competition-level mathematics. Examining case studies from four industries—electricity, finance, health care, and information—a recent Brookings Institute study[2] endorsed the strong potential of AI in driving long-term growth in productivity. From rising job postings requiring AI skills and increasing business investment in AI initiatives to AI integration in iPhone OS and self-driving cars, AI is undeniably part of how we work and live.

More interactions with AI are changing the public perception and attitude toward it. The *AI Index Report*[1] cited rising optimism toward AI globally (although unevenly across the geographic regions) and sharp increases in AI use in business functions by companies/organizations. Surveying more than 48,000 people from 47 countries across the globe, the Trust in AI Research program[3] found that while public AI adoption and usage have increased since 2022, trust in AI has declined and worry about its risk has increased. People in emerging economies appear to have higher optimism toward the benefits

Digital Object Identifier 10.1109/MITP.2025.3585226
Date of current version 15 August 2025.

of AI. In the United States, the 2025 EDUCAUSE AI Landscape Study[4] surveyed faculty and staff at higher education institutions and, in comparison to the 2024 survey, found increases in respondents' perceptions regarding AI's priority in an institution's strategic planning, policy setting, and faculty/staff professional development efforts. The reported institutional functions using AI range widely from administrative assistance to teaching assistance, and majorities of respondents cited academic integrity, curriculum, and assessment as areas of teaching and learning mostly impact by AI. Similarly, a series of interviews at 19 universities across the United States and Canada[5] suggest that the importance of AI is now widely recognized by faculty and staff, even among "nonadopters" or those unfamiliar with AI. Interestingly, both studies point out the disparity of support resources toward equitable and affordable access to AI, as well as AI literacy development for students, faculty, and staff.

One consensus among the studies cited above is the public demand and desire for more AI regulations. In response, AI governance efforts have ramped up, with governments developing frameworks as an attempt to safeguard AI data security, fairness, and transparency. In 2024, U.S. states passed more than 400 AI-related bills into law, with California leading the record with a total of 42 legislative bills on AI. For example, AB 2876 mandates the incorporation of AI literacy content into K-12 math, science, and history-social science curricula, as well as the use of AI literacy as part of the criteria for evaluating instructional materials. SB 942, the California AI Transparency Act, requires AI providers to make available an AI detection tool at no cost to the user, in addition to offering other means of disclosure to enhance transparency for AI-generated content.

The way we interact with AI, and consequently the policies and practices surrounding AI use, will continue to evolve as AI tools advance at a phenomenal speed. Futurepedia (https://www.futurepedia.io/ai-innovations#leaderboard) reported 4,458 AI innovations between 31 October 2024 and 19 May 2025.[6] New or improved tools, such as Gemini, GPT 4o, Sora, Firefly 3, Deepseek, Perplexity, and Nova—to just name a few—were released in the past year. AI availabilities are not adopted at the same speed in most higher education institutions, but the community would not be served well if colleges and universities do not invest to facilitate access to AI tools and support student, faculty, and staff's appropriate use of such tools.

AI STRATEGIES AT CALIFORNIA STATE UNIVERSITY, FULLERTON

The advancement of AI systems, the simultaneous excitement and anxiety among the community, the desire for more regulations and guidelines, and the need for more access and support for AI are equally felt at California State University, Fullerton (CSUF), a large public university in the United States. With more than 43,000 students and more than 4,000 faculty and staff, the university undertook a multipronged approach in 2024 to help a large organization navigate the rapidly changing world of AI.

The multipronged approach encompasses three foci: awareness, literacy development, and integration of AI into teaching and learning practices. To expand and enhance awareness of AI, multiple events on campus and in surrounding communities took place. Led by a team of faculty and staff who are AI "early adopters," campuswide AI conferences took place every semester to introduce the university community to new AI developments, broaden faculty and staff knowledge and understanding of AI, and provide a safe venue for stakeholders to discuss their opinions about (for or against) AI. Externally, championed by university senior leadership (e.g., the provost and vice president of Academic Affairs), presentations on AI topics were provided to key community partners, such as local district superintendents and legislators.

AI literacy development is coordinated through the Faculty Development Center (FDC, https://fdc.fullerton.edu/teaching/ai.html), the university hub for teaching and learning professional development. Faculty leaders who are well versed in AI systems developed a Canvas (learning management system) site that hosts a wealth of resources for faculty as they adapt to AI use in their teaching practices. Among these resources, faculty can start with an "AI crash course" as an on-ramp to AI integration, browse AI-infused teaching materials, and explore existing syllabi and assessment tools. While this Canvas site is provided via a self-enroll and self-pace model, the FDC also provides a wide range of workshops that provide interactive guidance to help build campus AI expertise. Participant efforts are also recognized via an AL Learning Trailblazer Certificate to incentive faculty and staff engagement.

Complementing the AI literacy development efforts, CSUF also invested in enabling equitable access to AI tools for all community members. In addition to existing tools—such as ChatGPT Edu, Copilot, Gemini, NotebookLM, and Adobe Firefly—TitanGPT (https://www.fullerton.edu/it/projects/titangpt/), an AI system powered by ChatGPT-4o, was made available to

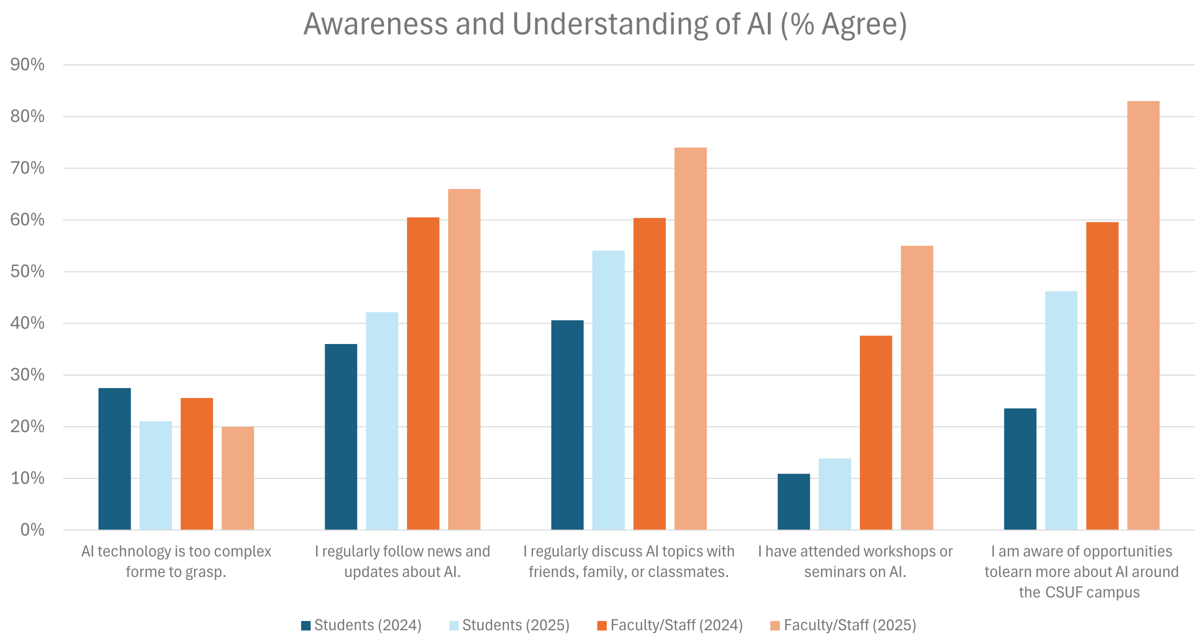


FIGURE 1. Survey responses for AI awareness and understanding.

all students, faculty, and staff in December 2024. This tool is designed to minimize the technology equity gap, protect user data in a secure environment, and integrates seamlessly with the university IT system. Since its launch, multiple AI projects have taken place, including faculty collaborating with instructional designers to integrate AI assignments and operational units developing specialized agents (e.g., advising, registration) to optimize business processes.

The integration of AI into teaching and learning at CSUF will soon be championed by an AI Center for Excellence devoted to leveraging AI to support curricular and research efforts to help students become workforce- and future-ready. Safeguarding all of these efforts are guidelines that speak to proper use of AI technology (https://fdc.fullerton.edu/teaching/using-ai-technology-guidelines.html) and encourage ethical use of AI (https://fdc.fullerton.edu/teaching/ai.html) across diverse academic contexts. As fast as AI systems evolve, these guidelines are also updated regularly to address emerging AI functionalities and accompanying concerns.

COMMUNITY PERCEPTIONS OF AI AT CSUF

To gauge the impact of the aforementioned AI strategies, a campuswide survey was administered to all students, faculty, and staff at CSUF for a two-week period in spring 2025. This survey serves as a follow-up to an identical survey administered in spring 2024[7] to capture any changes in campus community's perceptions of and experiences with AI. The survey consisted of five main sections and included questions that focused on people's awareness and understanding of AI (e.g., "I regularly follow news and updates about AI"), experience and usage of AI (e.g., "I use AI outside of my classwork"), attitudes and expectations toward AI (e.g., "I have concerns about AI's impact on job security"), education and training in AI (e.g., "I am actively seeking opportunities to learn more about AI"), and the AI tools they regularly use (e.g., ChatGPT). Most items follow a six-point Likert scale, seeking respondents' level of agreement with each statement from "strongly disagree" to "strongly agree." The spring 2024 survey resulted in more than 7,600 responses, including 6,488 student responses, while the spring 2025 survey collected more than 6,400 responses, 82% of which are from students. We interpret the differences observed between the two surveys as a proxy for the impact of AI strategies implemented on campus during the year.

All members of the campus community expressed significant increase in their awareness and understanding of AI (Figure 1). It appears that students and faculty/staff alike engaged more with opportunities to learn more about AI, and correspondingly, their concerns about AI complexity reduced slightly.

In terms of AI usage (Figure 2), it appears that the campus community is becoming more comfortable using AI tools in their studies and the value or benefits of these tools are becoming more recognized. This survey

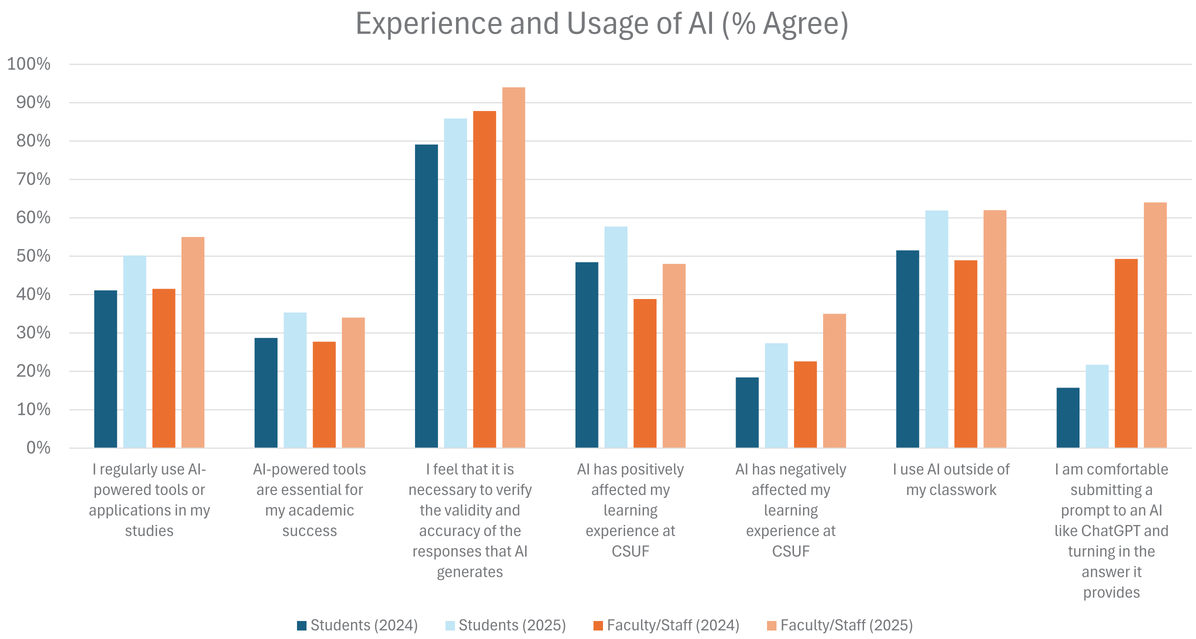


FIGURE 2. Survey responses for AI experience and usage.

result is corroborated by the observation that the number of on-campus AI queries nearly doubled from April 2024 (~82,000 queries) to April 2025 (~162,000 queries). However, just like the general public, students and faculty/staff both reported more concerns regarding the validity and accuracy of AI-generated responses. Interestingly, while more respondents reported agreement with the positive impact of AI on learning experiences, comparable increases are also seen for its negative impact. Perhaps as community members become more familiar with AI, their ability to evaluate it critically also improves, hence the more complex perceptions.

In comparison to a year ago, significantly more students and faculty/staff agreed that CSUF offers adequate AI training opportunities, serving as positive feedback for the institutional efforts detailed above (Figure 3). The interest in seeking formal training decreased slightly, perhaps an indication that more people felt their learning needs are met through all of the professional development opportunities. It appears that AI is perceived to be more integrated in the curriculum and coursework, although skepticism about AI remained high.

Despite the increased exposure and engagement with AI, the campus community's concerns about AI remain high (Figure 4). The worries regarding personal privacy, AI transparency, and impact on future job security is more heightened than before. The concern for the ethical use of AI remains for the majority of respondents.

Consistently, both students and faculty/staff remain worried about AI's impact on creativity and long-term society impact (Figure 5). The desire of regulating AI development to reduce risk remains strong, similar to the general population.

DISCUSSION

The observed survey response changes over the last year at CSUF echo the general public's perceptions and attitudes toward AI. The campus efforts to increase awareness, expertise, and integration of AI in teaching, learning, and operations appeared to have helped our students, faculty, and staff to become more accepting of AI's importance, more familiar with the AI concept, more engaged with AI tools, and more comfortable using the various tools. But at the same time, as they become more sophisticated users, they arguably become more critical of the uncertainty and potential risks associated with AI development. AI validity, accuracy, transparency, ethical use, and potential impact on future jobs are among the top issues the respondents expressed concerns for. The responses from students and faculty/staff are also more aligned in 2025 compared to a year ago, suggesting perhaps an institutional synergy.

We are encouraged by the survey results, as they suggest that our AI strategies are effective in helping our large university adapt to a future in which AI plays a significant role. The critical stance our community holds points out the needs for continued training and professional development that evolve with the AI

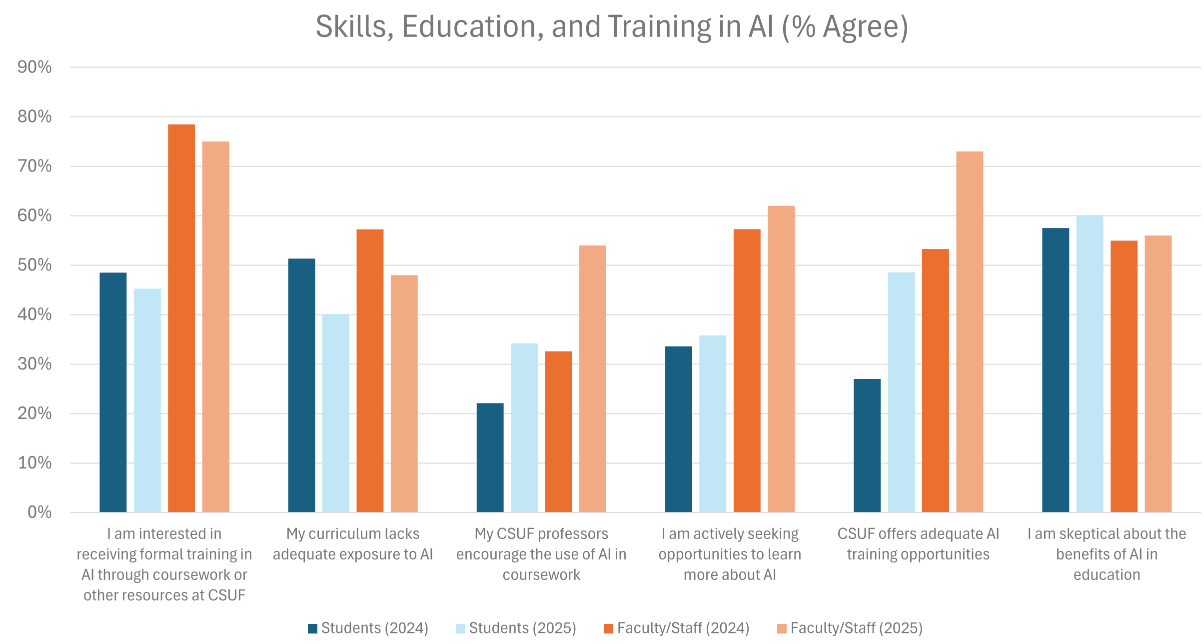


FIGURE 3. Survey responses for skills, education, and training in AI.

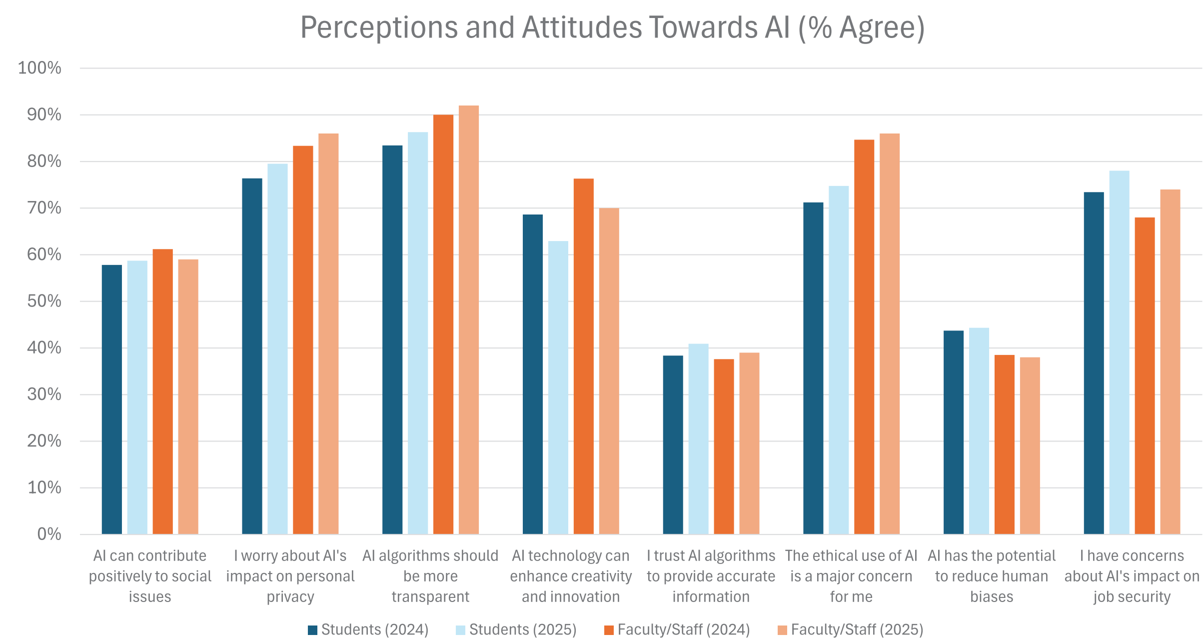


FIGURE 4. Survey responses for perceptions and attitudes toward AI.

systems. For the AI enthusiasts, we need to continue to strengthen institutional guidelines to avoid overzealous use of AI at the risk of impeding human creativity and academic integrity. For the AI nonadopters, we need to seek alternative ways to channel the skepticism into at least "guarded experimentation" of AI to adapt to today's student needs, workforce demand, and technological advancement.

It is not just AI that is rapidly evolving. People's learning needs are changing as well. As Arizona State University President Michael Crow[8] and Matter and Space (https://matterandspace.com/) founder Paul LeBlanc (former President of Southern New Hampshire University) pointed out, learning is becoming more personalized, flexible, and adaptable. "AI-powered higher education" is here. While AI is not (and may never be) perfect,

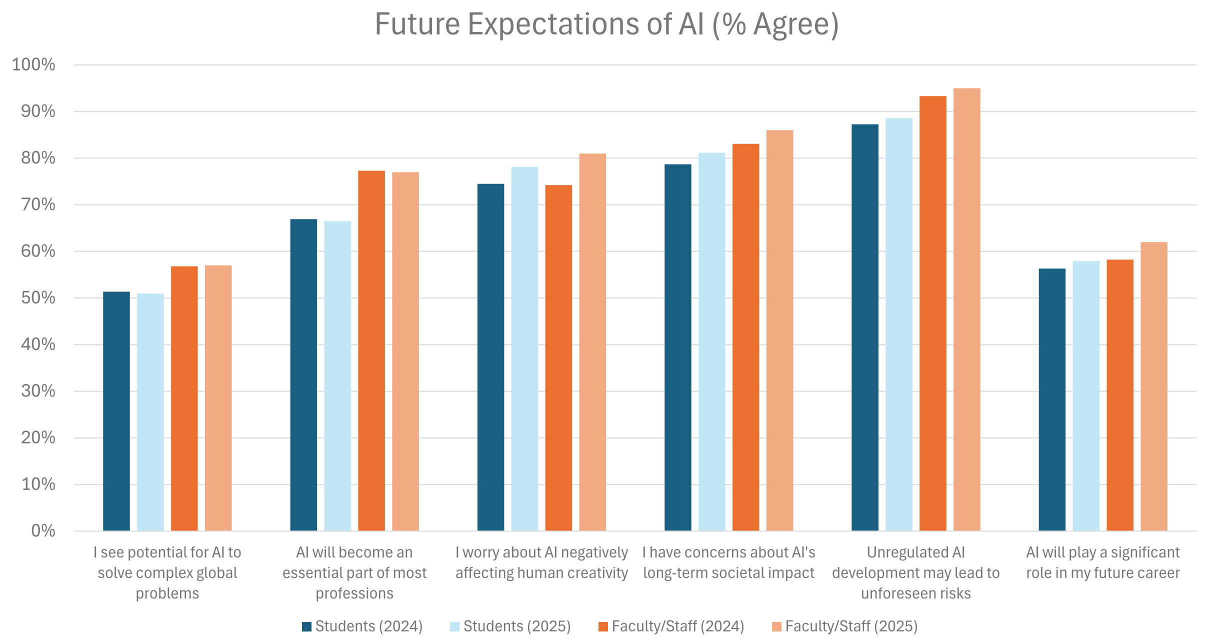


FIGURE 5. Survey responses for future expectations of AI.

it is absolutely necessary for higher education to incorporate AI with an open and agile mindset.

REFERENCES

1. N. Maslej et al., "The AI index 2025 annual report," Stanford University, Stanford, CA, Apr. 2025. [Online]. Available: https://hai.stanford.edu/ai-index/2025-ai-index-report
2. M. N. Baily and A. T. Kane, "Harnessing AI for economic growth: insights from electricity, finance, health care, and information sectors," *Brookings*, Apr. 8, 2025. [Online]. Available: https://www.brookings.edu/articles/harnessing-ai-for-economic-growth/
3. N. Gillespie, S. Lockey, T. Ward, A. Macdade, and G. Hassed, "Trust, attitudes and use of artificial intelligence: A global study 2025," The Univ. of Melbourne, Melbourne, VIC, Australia and KPMG, 2025. [Online]. Available: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2025/05/trust-attitudes-and-use-of-ai-global-report.pdf
4. J. Robert and R. M. McCormack, "2025 EDUCAUSE AI landscape study: Into the digital AI divide," EDUCAUSE, Boulder, CO, USA, Res. Rep., Feb. 2025. [Online]. Available: https://library.educause.edu/resources/2025/2/2025-educause-ai-landscape-study
5. C. Baytas and D. Ruediger, "Making AI generative for higher education: Adoption and challenges among instructors and researchers," *Ithaka S+R*, May 1, 2025. [Online]. Available: https://sr.ithaka.org/wp-content/uploads/2025/04/SR-Report-Making-AI-Generative-for-Higher-Ed-05012025.pdf
6. "The latest AI innovations," *Futurepedia*, May 19, 2025. [Online]. Available: https://www.futurepedia.io/ai-innovations
7. A. Dabirian and S. Swarat, "Artificial intelligence in higher education: Community perceptions at a large U.S. university," *IT Prof.*, vol. 26, no. 4, pp. 92–96, Jul./Aug. 2024, doi: 10.1109/MITP.2024.3434068.
8. B. Penprase, "'Evolve or die' – Michael Crow's challenge to US higher education," *Forbes*, May 1, 2025. [Online]. Available: https://www.forbes.com/sites/bryanpenprase/2025/05/01/evolve-or-die–michael-crows-challenge-to-us-higher-education/

AMIR DABIRIAN is provost and vice president for Academic Affairs and a professor of marketing with the Department of Marketing, California State University, Fullerton, Fullerton, CA, 92831, USA. Contact him at adabirian@fullerton.edu.

CHRISTOPHER SWARAT is the dean for the College of Professional and Continuing Education at California State University, Long Beach, Long Beach, CA, 90840. Contact him at chris.swarat@csulb.edu.

SU SWARAT is the senior associate vice president for Institutional Effectiveness and Planning at California State University, Fullerton, Fullerton, 92831, CA, USA. Contact her at sswarat@fullerton.edu.

EDITOR: Guansong Pang, gspang@smu.edu.sg

DEPARTMENT: AI COMMUNITY

This article originally appeared in Intelligent Systems vol. 39, no. 3, 2024

Foundation Models for Education: Promises and Prospects

Tianlong Xu and Richard Tong, *Squirrel Ai Learning, Bellevue, WA, 98004, USA*
Jing Liang, Xing Fan, and Haoyang Li, *Squirrel Ai Learning, Shanghai, 200233, China*
Qingsong Wen, *Squirrel Ai Learning, Bellevue, WA, 98004, USA*

With the advent of foundation models like ChatGPT, educators are excited about the transformative role that artificial intelligence (AI) might play in propelling the next education revolution. The developing speed and the profound impact of foundation models in various industries force us to think deeply about the changes they will make to education, a domain that is critically important for the future of humans. In this article, we discuss the strengths of foundation models, such as personalized learning, education inequality, and reasoning capabilities, as well as the development of agent architecture tailored for education, which integrates AI agents with pedagogical frameworks to create adaptive learning environments. Furthermore, we highlight the risks and opportunities of AI overreliance and creativity. Finally, we envision a future where foundation models in education harmonize human and AI capabilities, fostering a dynamic, inclusive, and adaptive educational ecosystem.

With the emergence of foundation models and generative AI (GenAI),[1] the implications for various science and technological domains have been rapidly explored to address real-world problems.[2,3] Similarly, the integration of foundation models with education has naturally arisen as a promising avenue,[4,5] particularly as large language models (LLMs)[6] are inherently instructive and can function like an extremely knowledgeable "teacher." Along with the trend, it is crucial to consider how to frame foundation models for education, leverage GenAI's unique advantages, and manage potential risks to traditional education. Foundation models, exemplified by ChatGPT, present a dual-edged sword in education, sparking debates over academic integrity versus innovative teaching aids. They offer the potential to assist students in studying and learning. However, concerns arise regarding the rigor of foundation models, as some early applications have shown limitations.[7] Therefore, finding a balance between utilizing AI's benefits and maintaining academic rigor is essential for the future of education[8]

Our major contributions are highlighting the strengths of foundation models in personalized learning, educational inequality, and reasoning capabilities, proposing an agent architecture for education, and at the same time, warning some risks of AI in education. Our major novelty is to establish a framework of future education foundational models, strengthening on educational penalization, being prepared for the overwhelming advancement of AI capabilities, and maintaining absolute human competitiveness in key capacities such as problem solving, critical thinking, and creativity.

STRENGTHS OF FOUNDATION MODELS FOR EDUCATION

Personalized Learning

The technical prowess of foundation models is revolutionizing education, as seen in the recent advancements in LLMs.[9] These advancements enhance the capabilities of LLMs to offer tailored feedback that considers students' personal preferences and historical interactions, dramatically improving individual

1541-1672 © 2024 IEEE
Digital Object Identifier 10.1109/MIS.2024.3398191
Date of current version 11 June 2024.

learning experiences. In practice, there are already some pioneers actively pursuing such directions. For instance, Khanmigo[10] utilizes LLMs to simulate the benefits of personal tutoring, functioning as a virtual writing coach that promotes critical thinking and problem solving. Squirrel AI[11] develops the large adaptive model encompassing foundation models, advanced retrieval augmented generation, and educational AI Agent, which can capture the intricate relationships between knowledge points, topics, and students' learning abilities for better personalized learning solutions. Duolingo's Duolingo Max[12] leverages LLMs for adaptive learning through roleplay, engaging users in lifelike conversations that seamlessly integrate into their learning paths. These platforms demonstrate how generative AI's nuanced understanding of context and personalized interaction can enhance education, making learning more responsive and interactive, much like a human tutor's guidance. Furthermore, industrial advancements underscore the need to build foundation models for education, which align with the principle of fostering individuals' holistic development and nurturing talents capable of innovation and independent thought in their fields.

Addressing Education Inequality

One widely recognized pain point for traditional education is educational inequity, whose root causes include resource allocation, teacher training, curriculum relevance, and social factors.[13] The introduction of foundation models can address these issues directly and also contribute to a broader strategy for achieving education equity. By leveraging its ability to process extensive data, foundation models can pinpoint the exact needs of diverse communities, ensuring resources are allocated more fairly. It enables personalized teacher training, breaking down geographical barriers, and uplifting educational quality across all regions. Additionally, adaptive learning technologies based on foundation models can tailor educational content to individual student backgrounds, making learning more engaging and accessible. This approach not only tackles the practical aspects of educational inequity but also combats social prejudices, fostering a more inclusive and equitable educational landscape.

Reasoning Capabilities

With the giant parameter space established during pretraining, LLMs have developed strong reasoning capabilities that continue to grow. In this domain, leveraging foundation models and adaptive learning techniques for math education is one of the most widely explored directions.[14,15] For example, many recent LLM-based works[16,17] have been tested or developed as the solver to the K12 level math problems, including arithmetic, geometry, equation sets, and their performance over some math word problem datasets like GSM8K are satisfactory. To further adopt LLMs for pedagogical purposes, the follow-up study explores the research questions on whether LLMs can correct students' wrong answers.[18] The step-by-step reasoning capabilities of Gemini[19] and many others have shown GenAI's strong potential in conquering sophisticated problems and positioning the "mistake steps" students might have in subjects including but actually will not be limited to mathematics. Such capabilities will be a strong addition to the teaching forces, which, in the one-on-one tutoring manner, significantly boost education effectiveness. Therefore, they need to be embedded as a strong backbone while creating the foundation models for education.

AGENT ARCHITECTURE FOR EDUCATION

To harness the potential of foundation models in the adaptive instructional environment, we foresee a new type of system architecture built on top of AI agents, as shown in Figure 1. This architecture can manage diverse and complex inputs for various pedagogical situations, adapt to changing contexts and curricula in a self-improvable manner,[20] and navigate the often ambiguous and interactive demands of students and educators.[21] It can be broken down into three components.

1) *Core agent architecture*: At the heart of the system are specialized agents responsible for distinct cognitive functions. These may range from diagnosis, forecasting, and problem-solving to providing psychological support. Each agent typically integrates both symbolic reasoning and neural network capabilities leveraging LLM and other foundation models.
2) *Agent orchestration and integration framework*: This layer serves as the "environment" that hosts the agents and enables their interaction, not only among themselves but also with external tools and platforms. A well-designed environment facilitates channel-based communication, where student behaviors, interactions, and other resources are funneled into a single session. This consolidated session offers real-time interaction and feedback mechanisms between agents, students, and educators. It also logs these activities for knowledge tracing, model refinement, or compliance monitoring.
3) *Pedagogical and ontological framework*: Beyond what is traditionally called a content management

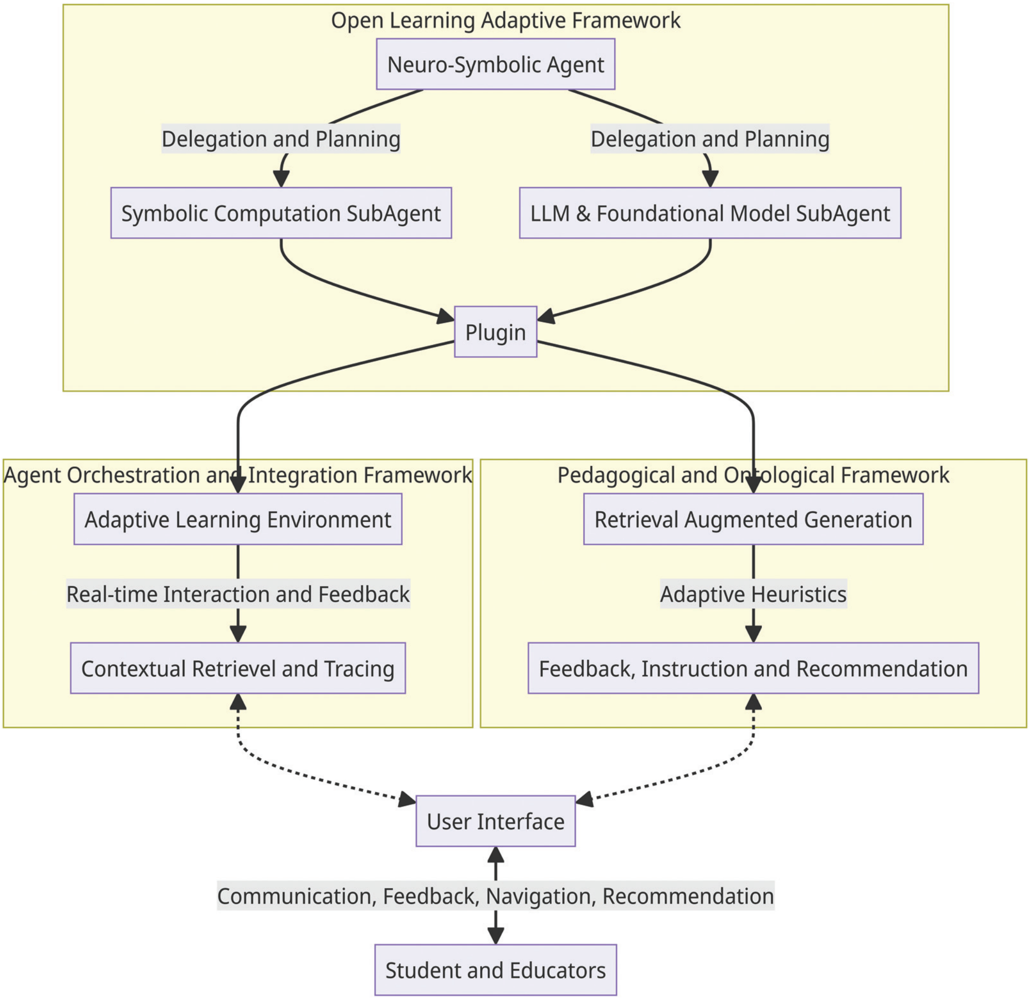


FIGURE 1. The agent architecture abstraction for education.

system, this framework operates at the intersection of content, learning objectives, and pedagogical strategies. It would likely be built upon an ontology framework that interlinks resources with learning goals and pedagogical heuristics.

RISKS AND POTENTIAL OPPORTUNITIES

Overreliance

Responsible AI has been a widely discussed topic since the introduction of AI. According to the latest framework Microsoft has proposed, there are six critical components: fairness, inclusiveness, reliability and safety, privacy and security, transparency, and accountability.[22] Beyond these, the potential tendency of overreliance as we continue to interact more with GenAI is worth being discussed as well. This issue mainly come from the concerns of AI implications on education; however, it could also extend to a wider range of impact given the "instructive" characteristics of most GenAI products. GenAI's ability to provide instant information might lead to a dependency that undermines critical thinking and the motivation for self-led learning. To combat this, AI should be used to encourage deeper inquiry, not just quick answers. Integrative teaching strategies that demand independent research and critical thinking are key to preventing overreliance on AI, ensuring students retain their ability to learn autonomously. This balance is crucial for the responsible incorporation of AI in education, presenting a challenge for educators and policymakers to foster independent, inquisitive learners while leveraging AI's advantages.

AI Creativity

A further step in thinking beyond the overreliance is whether foundation models can be truly innovative hinges on their learning capacity. Some argue that AI systems like GPT-4 learn from vast datasets and may exhibit creativity,[23] a view that could be implicitly supported by the "larger is better" hypothesis. However, the speculation remains open, as the extent to which AI can originate novel ideas is still unproven. Regardless of AI's potential for creativity, the emphasis in education should remain on nurturing human ingenuity. If AI is indeed capable of innovation, the challenge is to ensure it complements rather than competes with human creativity. By promoting educational frameworks that prioritize independent and critical thinking skills, we can ensure the dominance of human creativity, either using GenAI as a tool to enhance and amplify (rather than replace) the creative process, or using it as a virtuous competitor to maintain human competitiveness.

CONCLUSION AND FUTURE VISION

In the evolving landscape of foundation models, personalized learning emerges as a pivotal force in enriching educational experiences. It caters to the diverse needs, preferences, and abilities of each learner, thereby advancing educational equity. While technology serves to enhance these experiences, it is imperative that the essence and ultimate responsibility of decision-making remain firmly anchored in human hands. Looking forward, the envisioned educational foundation model heralds a shift towards a more dynamic, inclusive, and adaptive framework. This framework seeks to harmonize the strengths of human educators with the capabilities of GenAI technology, thereby preparing learners to navigate both foreseeable and unforeseen challenges with resilience and adaptability.

The future of education is envisioned as a realm where foundation models serve to amplify the value from human's potential and the vast amount of knowledge accumulated rather than a total replacement, such that AI and human's coevolution can progress towards an ideal direction. Future education is tailored to every individual's unique journey, empowering each learner to excel and realize full potential. This vision presents a holistic blueprint for cultivating educational environments that elevate human capacities, ensuring that learners from all backgrounds can thrive and maintain their utmost competitiveness in problem-solving capabilities, critical thinking, and creativity.

REFERENCES

1. Z. Epstein et al., "Art and the science of generative AI," *Science*, vol. 380, no. 6650, pp. 1110–1111, 2023, doi: 10.1126/science.adh4451.
2. M. Jin et al., "Large models for time series and spatio-temporal data: A survey and outlook," 2023, *arXiv:2310.10196*.
3. S. Pan, Y. Zheng, and Y. Liu, "Integrating graphs with large language models: Methods and prospects," *IEEE Intell. Syst.*, vol. 39, no. 1, pp. 64–68, Jan./Feb. 2024, doi: 10.1109/MIS.2023.3332242.
4. S. Wang et al., "Large language models for education: A survey and outlook," *arXiv Preprint arXiv:2403.18105*, 2024.
5. H. Li et al., "Bringing generative AI to adaptive learning in education," 2024, *arXiv:2402.14601*.
6. W. X. Zhao et al., "A survey of large language models," 2023, *arXiv:2303.18223*.
7. L. Huang et al., "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," 2023, *arXiv:2311.05232*.
8. "AI can transform education for the better," *The Economist*, Jan. 2024. [Online]. Available: https://www.economist.com/business/2024/01/11/ai-can-transform-education-for-the-better
9. P. Denny et al., "Generative AI for Education (GAIED): Advances, opportunities, and challenges," 2024, *arXiv:2402.01580*.
10. *Khan Academy*. Accessed: Mar. 2024. [Online]. Available: https://www.khanacademy.org/
11. *Squirrel AI*. Accessed: Mar. 2024. [Online]. Available: https://www.squirrelai.com
12. *Duolingo*. Accessed: Mar. 2024. [Online]. Available: https://www.duolingo.com/
13. W. Holmes, and K. Porayska-Pomsta, *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*. New York, NY, USA: Taylor & Francis, 2022.
14. H. Li, T. Xu, J. Tang, and Q. Wen, "Automate knowledge concept tagging on math questions with LLMs," 2024, *arXiv:2403.17281*.
15. Z. Liang, W. Yu, T. Rajpurohit, P. Clark, X. Zhang, and A. Kalyan, "Let GPT be a math tutor: Teaching math word problem solvers with customized exercise generation," in *Proc. Empirical Methods Natural Lang. Process.*, 2023, pp. 362–378.
16. K. Wang et al., "Mathcoder: Seamless code integration in LLMs for enhanced mathematical reasoning," in *Proc. Int. Conf. Learn. Representations*, 2024.
17. X. Zhao, Y. Xie, K. Kawaguchi, J. He, and Q. Xie, "Automatic model selection with large language models for reasoning," in *Proc. Findings Empirical Methods Natural Lang. Process.*, 2023.
18. A.-Z. Yen, and W.-L. Hsu, "Three questions concerning the use of large language models to facilitate mathematics learning," in *Proc. Findings Empirical Methods Natural Lang. Process.*, 2023, pp. 3055–3069.
19. "Introducing Gemini: Our largest and most capable AI model," Google, Menlo Park, CA, USA, 2023.

Accessed: Feb. 8, 2024. [Online]. Available: https://blog.google/technology/ai/google-gemini-ai/

20. R. Tong, J. Rowe, and B. Goldberg, "Chapter 3 – Architecture implications for building macro and micro level self-improving AISs," in *Design Recommendations for Intelligent Tutoring Systems: Volume 7 - Self-Improving System*, A. M. Sinatra, A. C. Graesser, X. Hu, K. Brawner, and V. Rus, Eds., Orlando, FL, USA: US Army Research Lab., Human Research & Engineering Directorate, 2019, pp. 17–27. [Online]. Available: https://gifttutoring.org/attachments/download/3410/DesignRecommendationsforITS_Volume7_SelfImprovingSystemsBook.pdf
21. R. Tong, and T. X. Lee, "Trustworthy AI that engages humans as partners in teaching and learning," *Computer*, vol. 56, no. 5, pp. 62–73, May 2023, doi: 10.1109/MC.2023.3234517.
22. "What is responsible AI?" Microsoft, Redmond, WA, USA, 2024. Accessed: Feb. 8, 2024. [Online]. Available https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2
23. H. Wang et al., "Can AI be as creative as humans?" 2024, *arXiv:2401.01623*.

TIANLONG XU is a staff applied scientist at Squirrel Ai Learning, Bellevue, WA, 98004, USA. Contact him at tianlongxu@squirrelai.com.

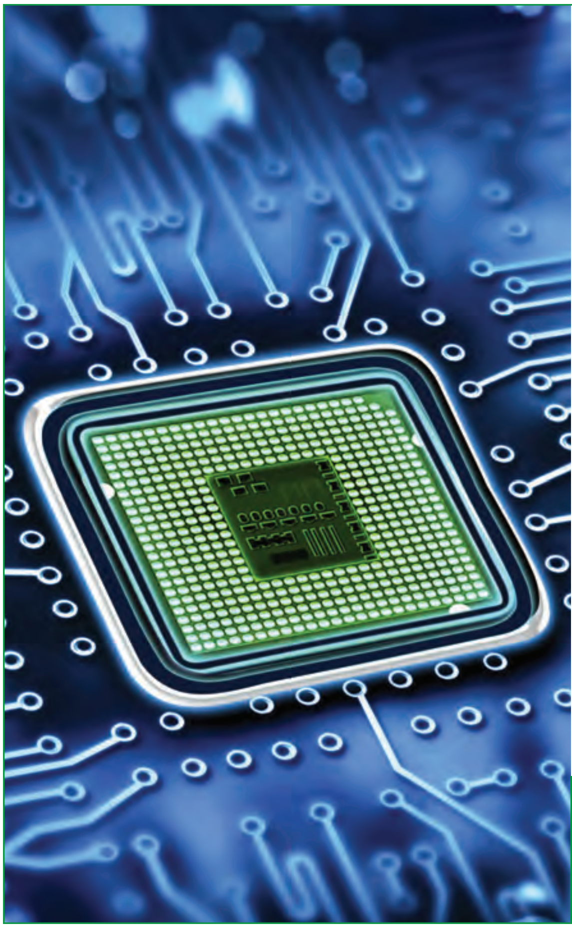
RICHARD TONG is a chief architect at Squirrel Ai Learning, Bellevue, WA, 98004, USA. Contact him at richard.tong@ieee.org.

JING LIANG is a cofounder of Squirrel Ai Learning, Shanghai, 200233, China. Contact her at joleenliang@squirrelai.com.

XING FAN is a CTO at Squirrel Ai Learning, Shanghai, 200233, China. Contact him at fanxing@songshuai.com.

HAOYANG LI is a founder and CEO at Squirrel Ai Learning, Shanghai, 200233, China. Contact him at derek@squirrelai.com.

QINGSONG WEN is a chief AI scientist at Squirrel Ai Learning, Bellevue, WA, 98004, USA. Contact him at qingsongwen@squirrelai.com.



Conference Calendar

IEEE Computer Society conferences are valuable forums for learning on broad and dynamically shifting topics from within the computing profession. With over 200 conferences featuring leading experts and thought leaders, we have an event that is right for you. Questions? Contact conferences@computer.org.

FEBRUARY

2 February

- AIxDKE (Int'l Conf. on AI x Data and Knowledge Eng.), Laguna Hills, USA
- BigComp (IEEE Int'l Conf. on Big Data and Smart Computing), Guangzhou, China
- ICSC (Int'l Conf. on Semantic Computing), Laguna Hills, USA

16 February

- ICNC (Int'l Conf. on Computing, Networking and Communications), Maui, USA

MARCH

6 March

- WACV (IEEE/CVF Winter Conf. on Applications of Computer Vision), Tucson, USA

16 March

- PerCom (IEEE Int'l Conf. on Pervasive Computing and Communications), Pisa, Italy

17 March

- SANER (IEEE Int'l Conf. on Software Analysis, Evolution and Reengineering), Limassol, Cyprus

20 March

- 3DV (Int'l Conf. on 3D Vision), Vancouver, Canada

21 March

- VR (IEEE Conf. on Virtual Reality and 3D User Interfaces), Daegu, Korea

22 March

- SSIAI (IEEE Southwest Symposium on Image Analysis and Interpretation), Santa Fe, USA

23 March

- SaTML (IEEE Conf. on Secure and Trustworthy Machine Learning), Munich, Germany

APRIL

12 April

- AST (IEEE/ACM Int'l Conf. on Automation of Software Test), Rio de Janeiro, Brazil
- FormaliSE (IEEE/ACM Int'l Conf. on Formal Methods in Software Eng.), Rio de Janeiro, Brazil
- ICSE (IEEE/ACM Int'l Conf. on Software Eng.), Rio de Janeiro, Brazil
- MOBILESoft (IEEE/ACM Int'l Conf. on Mobile Software Eng. and Systems), Rio de Janeiro, Brazil
- MSR (IEEE/ACM Int'l Conf. on Mining Software Repositories), Rio de Janeiro, Brazil

15 April

- COOL CHIPS (IEEE Symposium on Low-Power and High-Speed Chips and Systems), Tokyo, Japan

20 April

- DATE (Design, Automation & Test in Europe Conf.), Verona, Italy
- PacificVis (IEEE Pacific Visualization Conf.), Sydney, Australia

26 April

- ISPASS (IEEE Int'l Symposium on Performance Analysis of Systems and Software), Seoul, Korea

MAY

4 May

- HOST (IEEE Int'l Symposium on Hardware Oriented Security and Trust), Washington, DC, USA
- MOST (IEEE Int'l Conf. on Mobility, Operations, Services and Technologies), Detroit, USA

8 May

- BigDataSecurity (IEEE Conf. on Big Data Security on Cloud), New York City, USA
- CAI (IEEE Int'l Conf. on Artificial Intelligence), Granada, Spain
- HPSC (IEEE Int'l Conf. on High Performance and Smart Computing), New York City, USA
- IDS (IEEE Int'l Conf. on Intelligent Data and Security), New York City, USA
- SmartCloud (IEEE Int'l Conf. on

- Smart Cloud), New York City, USA

11 May
- SenSys (ACM/IEEE Int'l Conf. on Embedded Artificial Intelligence and Sensing Systems), Saint Malo, France

12 May
- RTAS (IEEE Real-Time and Embedded Technology and Applications Symposium), Saint Malo, France

13 May
- FCCM (IEEE Annual Int'l Symposium on Field-Programmable Custom Computing Machines), Atlanta, USA

15 May
- ICES (Int'l Conf. on Energy Storage), Shenyang, China

18 May
- CCGrid (IEEE Int'l Symposium on Cluster, Cloud and Internet Computing), Sydney, Australia
- ICFEC (IEEE Int'l Conf. on Fog and Edge Computing), Sydney, Australia
- ICST (IEEE Int'l Conf. on Software Testing, Verification and Validation), Daejeon, Korea
- S&P (IEEE Symposium on Security and Privacy), San Francisco, USA

19 May
- ICDE (IEEE Int'l Conf. on Data Eng.), Hong Kong, China
- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic), Sendai, Japan

25 May
- FG (IEEE Int'l Conf. on Automatic Face and Gesture Recognition), Kyoto, Japan
- IPDPS (IEEE Int'l Parallel and Distributed Processing Symposium), New Orleans, USA

JUNE

1 June
- ICHI (IEEE Int'l Conf. on Healthcare Informatics), Minneapolis, USA

3 June
- CBMS (IEEE Int'l Symposium on Computer-Based Medical Systems), Limassol, Cyprus

6 June
- CVPR (IEEE/CVF Conf. on Computer Vision and Pattern Recognition), Denver, USA

10 June
- SVCC (Silicon Valley Cybersecurity Conf.), San Jose, USA

22 June
- DCOSS-IoT (Int'l Conf. on Distributed Computing in Smart Systems and the Internet of Things), Reykjavik, Iceland
- ICSA (IEEE Int'l Conf. on Software Architecture), Amsterdam, Netherlands

26 June
- IEEE Cloud Summit, Washington, DC, USA

27 June
- ISCA (ACM/IEEE Annual Int'l Symposium on Computer Architecture), Raleigh, USA

JULY

6 July
- EuroS&P (IEEE European Symposium on Security and Privacy), Lisbon, Portugal
- ICALT (IEEE Int'l Conf. on Advanced Learning Technologies), Hung Yen, Vietnam

7 July
- COMPSAC (IEEE Annual Computers, Software, and Applications Conf.), Madrid, Spain
- ISVLSI (IEEE Computer Society Annual Symposium on VLSI), Kolkata, India