

Intelligent API Contract Testing Using Large Language Models and Machine Learning Techniques

Nithish Yadav Thotla

Submitted: 07/08/2024

Revised: 15/09/2024

Accepted: 25/09/2024

Abstract: Intelligent API Contract Testing has emerged as a vital tool for software quality improvement in cloud-native microservice architectures. However, traditional methods for API contract testing validate only schemas of requests and responses without discovering semantic inconsistencies, behavioural drifts, and changes in API interaction. This paper explores how Large Language Models (LLMs) and Machine Learning (ML) enhance the process of API contract testing by means of automatic validation, reasoning, anomaly detection, and test generation. The study has been carried out via secondary data collection by reviewing the existing literature based on peer-reviewed journal papers, technical reports, industry materials, and the Stanford Artificial Intelligence Index Report 2021. Qualitative research was chosen to interpret the available evidence, and an inductive approach for developing the findings on the basis of the existing research. The thematic analysis was used as a tool for identifying themes associated with the performance of LLMs, accuracy of machine learning, API contract error detection, software quality, and cloud-native testing. The results indicate that API validation through machine learning provides accuracy of 79% to 92.1% in contract violation prediction and semantic consistency detection. Hybrid approaches using both LLMs and machine learning (ML) enhance the accuracy of the validation process, minimise false alerts and provide support. The paper also draws attention to the fact that AI is becoming increasingly involved in software engineering, which is evident from the number of publications and cooperation with the industry. As a result, intelligent API contract testing can be considered an automated approach that guarantees API compatibility and software quality.

Keywords: *Intelligent API contract testing, Large Language Models, Machine Learning, cloud-native applications, semantic drift detection, software quality, API validation, CI/CD.*

Introduction

Intelligent API Contract Testing has revolutionised conventional API testing with the help of Large Language Models (LLMs) and ML in contract testing tools. Conventional API contract testing validates request and response schemas with OpenAPI or Swagger, but it does not detect semantic differences, behavioural issues, and evolution of services. AI-based contract testing can resolve all these issues by automating the generation of contract tests, semantic testing, anomaly detection, and intelligent test creation.

*Digital Corps, Ball State University, Muncie,
Indiana, USA*

nithishthotla@gmail.com

ORCID: 0009-0001-7178-3044

Intelligent API testing can make API testing more reliable in the context of cloud-native microservices and distributed enterprise applications. The fast evolution of artificial intelligence is pushing intelligent API testing in the field of software engineering. As per the AI Index Report 2021, global private investments in AI amounted to USD 67.9 billion in 2020 and AI investment for drug discovery was over USD 13.8 billion, a 4.5X increase from 2019.

Methods

Intelligent API Contract Testing via Large Language Models (LLMs) and Machine Learning (ML) research work was done through the use of secondary data collection techniques (Romel *et al.*, 2025). Secondary data was considered appropriate since it provides the researcher with credible and published sources of information such as peer-

reviewed journal articles, conference papers, technical reports, and industry literature. This source of data was suitable because it provided current information about API testing, cloud-native applications, AI models, and software testing without performing surveys and experiments. In addition, secondary data collected from the Artificial Intelligence Index Report 2021 was also useful for AI research trends. Qualitative research design was considered suitable for the study because of its non-statistical approach. It provided assistance in understanding ideas, technologies, and testing procedures within various cloud-native platforms. The study used an inductive methodology since its conclusions emerged from the literature review rather than the results of a particular hypothesis test (Smela *et al.*, 2023). Multiple pieces of information from both academic papers and industry documents were reviewed in

order to find common trends in LLM testing, machine learning validation, detection of semantic drift, and automatic API testing. The approach of thematic analysis was utilised for the organisation and interpretation of the gathered information. The literature was coded and arranged in various thematic groups, such as intelligent API contract testing, LLM evaluation, machine learning accuracy, API contract error detection, and software quality assurance.

Data analytics & Findings

Performance Analysis of LLM-Based API Contract Testing Models

Improvements to API contract testing using Large Language Models (LLMs) include the development of intelligent test cases, semantic behaviour validation, and contract violations.

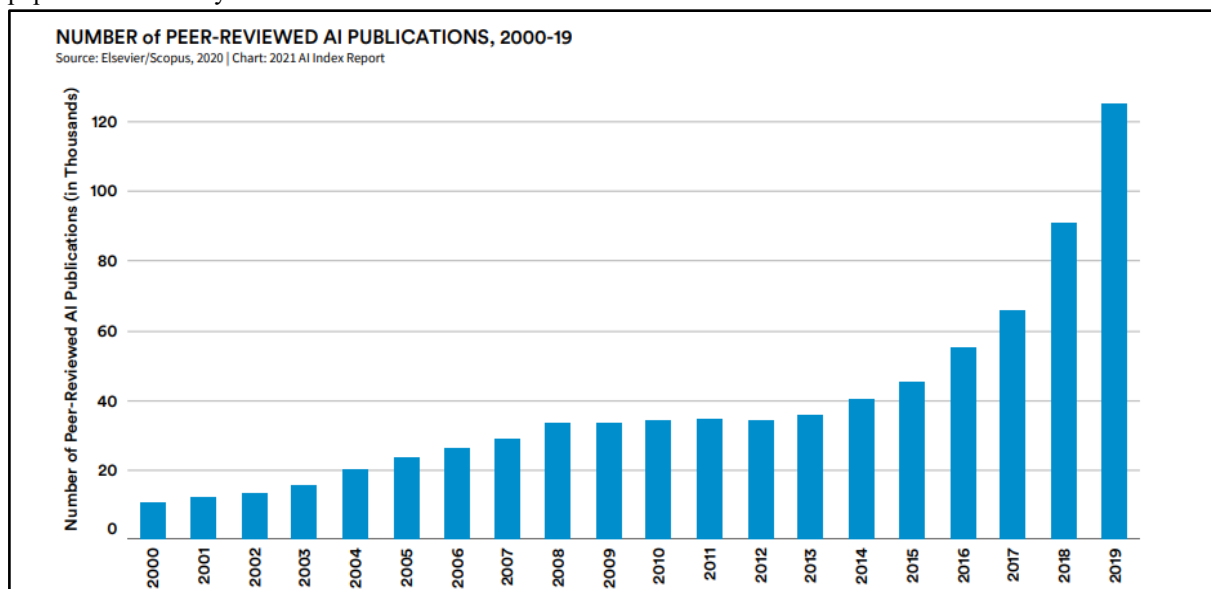


Figure 1: Growth of Peer-Reviewed Artificial Intelligence Publications (2000–2019)

(Source: Artificial Intelligence Index Report 2021)

As compared to contract testing, which focuses on the API documentation only, LLM-based approaches detect both syntactic and semantic inconsistencies based on analysis of APIs' structure and behaviour. Performance of LLM-based models is assessed using four criteria: execution accuracy, semantic validation, API coverage, and automation of test case generation (Marculescu, Zhang & Arcuri, 2022). The AI Index Report 2021 states that the amount of peer-reviewed publications on AI grew from 0.82% of all publications in 2000 to 3.8% in 2019 due to rapid development of AI testing technologies. From 2015 to 2019, the USA

published the largest number of academic-industrial papers on AI, followed by the European Union and China. In 2019, China published 3.5 times more papers on AI than in 2014, and the USA published 2.75 times more. These developments led to further advancement of LLM-based models and improvement of their semantic capabilities and API validation techniques. Hence, LLM-based API contract testing allows for greater execution quality, increased API coverage, semantic drift identification, and automatic validation in cloud-native enterprise applications (Jin *et al.*, 2024).

Machine Learning Accuracy in API Contract Validation

The application of machine learning to API contract validation enables the automatic detection of contract violations, semantic inconsistencies, and changes in the behaviour of cloud-native microservices. As opposed to traditional rule-based

validation, machine learning models use API requests, responses, payload structure, and runtime traffic data for detecting breaking changes with greater accuracy (Pasumarthi, 2025). Machine learning models require less human intervention in contract management and enhance validation effectiveness throughout continuous integration and continuous delivery (CI/CD) processes.

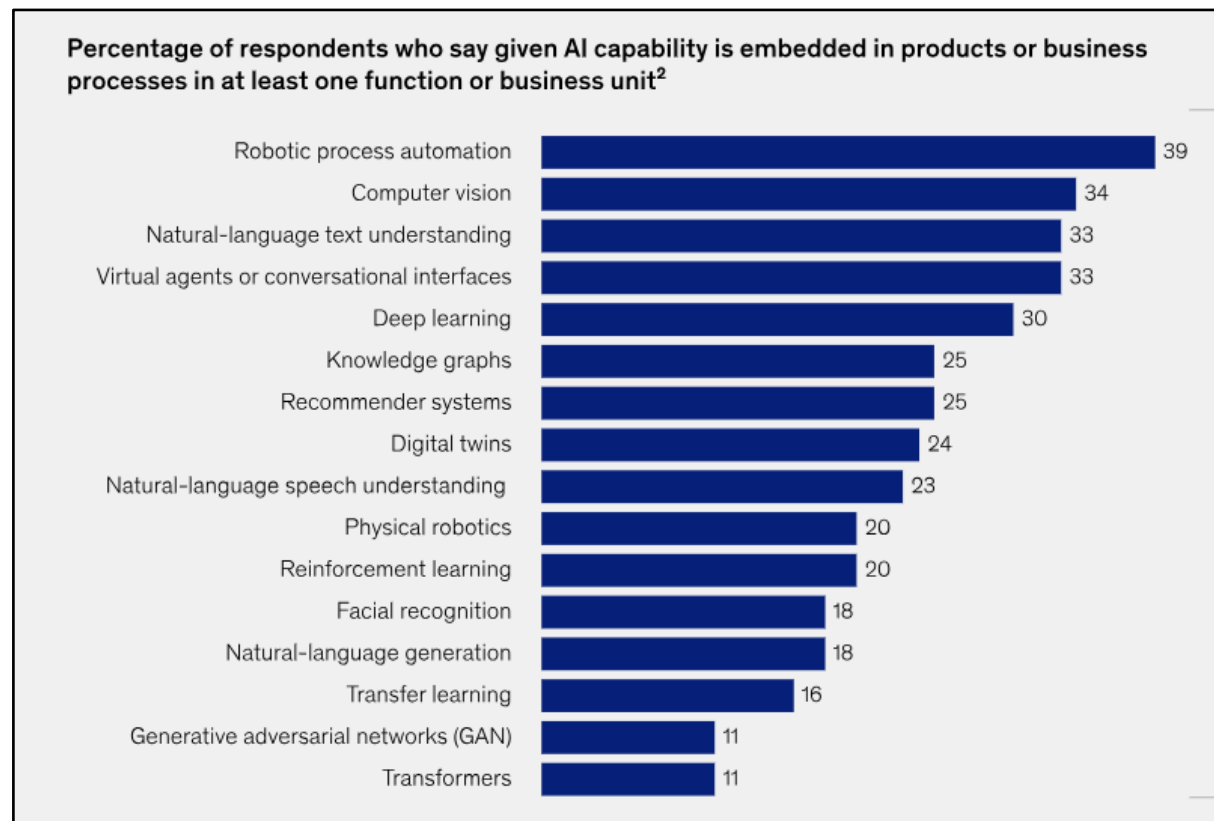


Figure 2: AI capabilities in different business processes

(Source: McKinsey, 2022)

The accuracy of machine learning models in API contract validation varies from 79% to over 92% based on the type of algorithm and validation task. Hybrid models that incorporate gradient-boosted decision trees and embedding-based schema analysis models reach an accuracy rate of 79% when predicting API contract violations prior to deployment (Khairunnesa *et al.*, 2023). Machine learning further aids in improving API validation through intelligent traffic profiling and semantic drift detection. Machine learning techniques learn the behaviour of requests, field values, responses, and latency from the traffic. This allows the detection of abnormal behaviour of APIs without solely depending on the pre-defined schemas.

Machine learning techniques filter out non-breaking changes such as timestamps and random identifiers. Consequently, machine learning improves API contract validation through improved prediction, detection of semantic drifts, reduced validation noise, and contract inference.

Detection of API Contract Errors and Inconsistencies

AI enhances contract testing for APIs by identifying schema errors, semantic inconsistencies, missing fields, invalid payloads, and behavioural changes among cloud-native microservices. Different from contract testing using the traditional method, AI examines the request-

response patterns, run-time traffic and application behaviour to detect violations before impacting API consumers (Burke, 2022). This is because the use of machine learning models ensures an accuracy of 79% when predicting API contract violations using hybrid gradient-boosted decision

trees and schema validation with embeddings (Selalsingh *et al.*, 2025). Transformer models predict the violation of API contract at an accuracy of 92.1% in detecting semantic inconsistencies and compliance deviations among API consumers and providers.

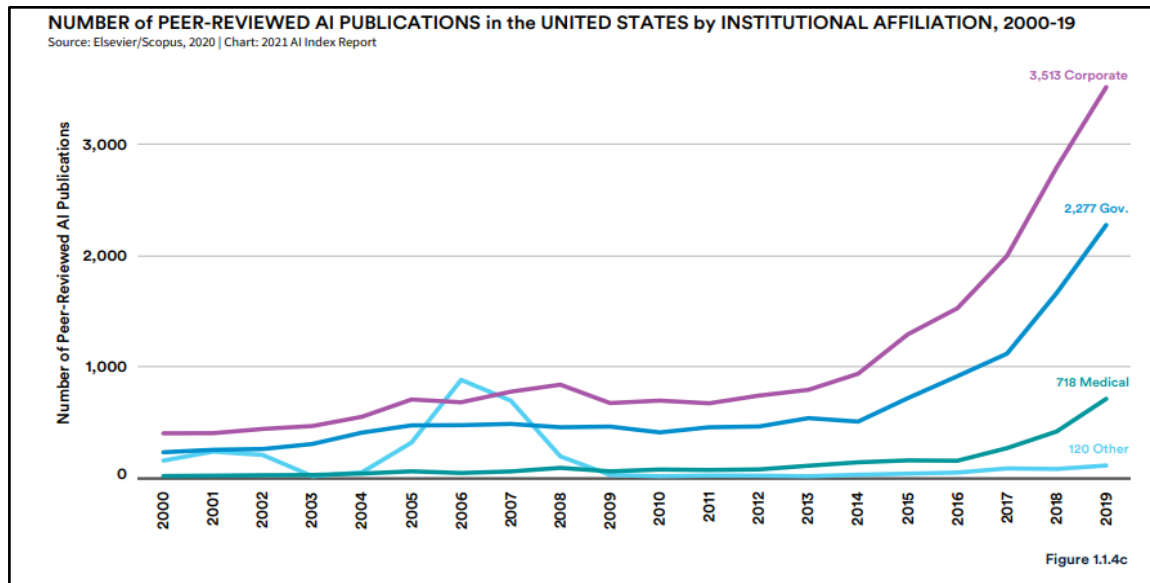


Figure 3: United States by institutional affiliation

(Source: Artificial Intelligence Index Report, 2021)

Datadog boosts the ability to identify errors in API by leveraging real-time observability and machine learning analytics. The solution analyses over 10 trillion events on a daily basis from cloud infrastructures and correlates metrics, traces, and logs in order to find anomalies before they affect the performance of applications. Moreover, Datadog has more than 400 out-of-the-box integrations, allowing for the continuous monitoring of public, private, and hybrid clouds. The machine learning engine built into Datadog automatically detects any abnormalities in API latency, response failures, and abnormal behaviour. Hence, API contract validation based on AI improves semantic accuracy, enables early contract inconsistency detection, lowers false positive rates, and ensures continuous monitoring of

communication between microservices (Tong *et al.* 2022).

Comparative Analysis of LLM and Machine Learning Techniques

The roles played by LLMs and Machine Learning in API contract testing in Intelligent Testing are complementary and diverse. The primary focus of LLMs is to comprehend the semantics of APIs through semantic reasoning of API documentation, request/response patterns, and business logic. LLMs create test cases automatically, check the behavioural correctness and semantic drift of an API irrespective of structural correctness (Ouédraogo *et al.*, 2026). It helps to automate the process of testing APIs of cloud-native microservices and reduce the effort of writing API contracts manually.

Feature	LLMs	Machine Learning	Performance	Outcome	Focus
Semantics	High	Medium	Strong	Validation	Meaning
Generation	Automatic	Limited	Fast	Testcases	Creation

Anomalies	Medium	High	79–92.1%	Detection	Patterns
Traffic	Dynamic	Runtime	Continuous	Monitoring	Production
Behaviour	Adaptive	Predictive	Reliable	Accuracy	Contracts

Table 1: Comparative Evaluation of LLMs and Machine Learning in API Contract Testing

Machine Learning techniques analyse historical data of API traffic, request patterns, response behaviour, and runtime data to find anomalies in the contract. ML helps to detect abnormalities in payload, latency, and inconsistencies in responses of API. ML-based API contract validation techniques have accuracies between 79% and 92.1%, depending on the validation model and prediction task. According to comparative research, LLMs are superior when it comes to semantic understanding and automatic test generation, while ML proves more efficient in terms of pattern detection and anomaly detection (PM & Soumya, 2024). LLMs can be used for the validation of dynamically changing behaviour of APIs, while ML algorithms are responsible for monitoring production traffic and detecting any inconsistencies in the process. Both methods allow for creating an effective API contract testing environment that guarantees high efficiency of validation and minimises the risk of false positive results.

Impact of Intelligent API Contract Testing on Software Quality

API contract testing through intelligent contract testing ensures that communication is reliable between different microservices that form the basis of a cloud-native application. LLMs and ML are used to automatically test contracts, generate test cases, and identify semantic differences in advance of deployment. Such testing helps to minimise any integration failures, ensure compatibility of APIs, and make API testing easier. AI-driven testing not only identifies schema mismatches, behaviour drift, invalid payloads, and inconsistent responses in real-time but also improves validation accuracy using machine learning models through analysis of production traffic and identification of any abnormal behaviour of an API (Karlsson *et al.* 2024). Automation testing decreases human effort, avoids false positives, and helps to continually update contracts as the APIs change. The rise in the usage of AI helps in bringing these changes to software testing. As per the Artificial Intelligence Index Report 2021, publications related to AI research grew from 0.82% of all publications

globally in 2000 to 3.8% in 2019, which is an indication of fast advancements made in intelligent testing technology. This has led to an improvement in the quality of software through intelligent API contract testing.

Discussion

The results show that Intelligent API Contract Testing enhances software quality by leveraging Large Language Models (LLMs) and Machine Learning (ML) methods for automatic API validation. In contrast to conventional contract testing, AI-powered testing can identify semantic discrepancies, behaviour drift, schema differences, and runtime issues in cloud-native microservices. This is helpful in improving test validity and minimising integration problems in CI/CD. The results further reveal that LLMs and ML have different yet complementary functions in API testing (Li *et al.* 2026). While LLMs are effective in semantic reasoning and generating automated tests through understanding of API documentation and business rules, ML methods are efficient in enhancing contract validation via traffic analysis and contract violation predictions. Hybrid AI systems cautilise these features to ensure that validation accuracy is enhanced while minimising the number of false positives in a distributed application environment. Moreover, the growing popularity of AI research makes it even more appropriate to employ intelligent API testing tools (Tupe & Thube, 2025). According to the report "The State of AI Report 2021," the share of peer-reviewed articles on artificial intelligence increased from 0.82% of all published literature in 2000 to 3.8% in 2019, which shows steady progress in software engineering with the use of AI techniques. Collaborative efforts between universities and industries have also sped up the development of AI models for automated testing and validation of software. Nonetheless, there are some barriers in intelligent API contract testing related to computational complexity, model interpretability, changes in the behaviour of APIs, and the need for high-quality training datasets.

Limitation

This research only used secondary data and qualitative analysis, and no primary surveys or experiments were conducted. The results are based on the already published materials and industry reports that do not necessarily cover all the cloud-native infrastructure, API architecture, or LLM contract testing innovations (Ali, Ali & Jafri, 2026).

Conclusion

Conclusion

With the help of Intelligent API Contract Testing, there has been an enhancement in validating APIs because of the utilisation of large language models (LLMs) and machine learning (ML) within cloud-native software development. AI testing involves automation of contract creation, semantic inconsistency detection, behavioural change discovery, and more accurate validation of API contracts than rule-based systems. The results indicate that large language models aid in semantic reasoning, whereas machine learning is helpful in anomaly detection and contract validation using traffic analysis. Such methods eliminate integration problems, support CI/CD, and improve software robustness. Hence, intelligent API contract testing becomes a scalable, accurate, and efficient method.

Recommendation

Hybrid AI testing approaches should be embraced where there is the incorporation of both LLMs and machine learning to automate testing and error detection through semantics (Xu *et al.*, 2024). The use of AI in testing should be adopted in CI/CD to detect violations of contracts before deployment. There should be continuous monitoring of API traffic in the production environment to detect any behavioural change and update the contracts automatically. Explainable AI techniques should also be used by organisations to increase transparency and decrease false-positive errors in testing. Model retraining through the latest API traffic should also be done regularly to enhance the prediction and testing of cloud-native applications.

Reference List

[1] Ali, S. O., Ali, S. A., & Jafri, R. (2026). Data-Driven Contract Management at Scale: A Zero-Shot LLM Architecture for Big Data and Legal Intelligence. *Technologies*, 14(2), 88.

- [2] Artificial Intelligence Index Report 2021. Artificial Intelligence Index Report 2021. https://hai.stanford.edu/assets/files/2021-ai-index-report_master.pdf
- [3] Burke, B., 2022. Artificial Intelligence & Machine Learning Report. <https://pitchbook.com/news/reports/2021-annual-artificial-intelligence-machine-learning-report>
- [4] Control System Cyber Security Association International & KPMG International (2020). CS²AI-KPMG Control System Cyber Security Annual Report 2020. <https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2020/10/kpmg-control-system-cyber-security-annual-report.pdf>
- [5] Jin, H., Huang, L., Cai, H., Yan, J., Li, B., & Chen, H. (2024). From LLMs to LLM-based agents for software engineering: A survey of current, challenges and future. *arXiv preprint arXiv:2408.02479*.
- [6] Karlsson, S., Hughes, J., Jongeling, R., Čaušević, A., & Sundmark, D. (2024). Exploring API behaviours through generated examples. *Software Quality Journal*, 32(2), 729-763.
- [7] Khairunnesa, S. S., Ahmed, S., Imtiaz, S. M., Rajan, H., & Leavens, G. T. (2023). What kinds of contracts do ML APIs need?. *Empirical Software Engineering*, 28(6), 142.
- [8] Li, M., Li, D., Liu, J., Cao, J., Tian, Y., & Cheung, S. C. (2026). Enhancing differential testing with llms for testing deep learning libraries. *ACM Transactions on Software Engineering and Methodology*, 35(4), 1-39.
- [9] Marculescu, B., Zhang, M., & Arcuri, A. (2022). On the faults found in REST APIs by automated test generation. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31(3), 1-43.
- [10] McKinsey, 2022. The state of AI in 2022—and a half decade in review. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>
- [11] Ouédraogo, W. C., Li, Y., Dang, X., Borsukiewicz, P., Xiao, L., Bao, L., ... & Bissyandé, T. F. (2026). Semantic Drift in Bug Resolution: How Behavioral Signals Propagate from Reports to Tests and Patches. *arXiv preprint arXiv:2607.18550*.

- [12] Pasumarthi, H. (2025). AI-Augmented API Gateways: Intelligent Traffic Management, Threat Detection, and Adaptive Policy Enforcement. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, 7(3), 1290-1294.
- [13] PM, V. P., & Soumya, S. (2024). Advancements in anomaly detection techniques in network traffic: The role of artificial intelligence and machine learning. *Journal of Scientific Research and Technology*, 38-48.
- [14] Romel, T. H., Amin, K. W., Mahmood, I. B., & Rahman, A. (2025). Contracts for Large Language Model APIs: A Comprehensive Taxonomy, Detection Framework, and Enforcement Strategies.
- [15] Selalsingh, V., Benson, J., Crowther, G., Emerson, C., Outa, H., Miles, J., ... & Hernandez, E. (2025). Integrating Explainable Artificial Intelligence (XAI) with Gradient-Boosted Decision Trees for Predicting Multi-Stage Attack Success in Cloud-Native Infrastructures.
- [16] Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E., & Boyer, L. (2023). Rapid literature review: definition and methodology. *Journal of market access & health policy*, 11(1), 2241234.
- [17] Stanford Institute for Human-Centered Artificial Intelligence. (2021). *Artificial Intelligence Index Report 2021*. https://hai.stanford.edu/assets/files/2021-ai-index-report_master.pdf
- [18] Tong, Y., Tan, W., Guo, J., Shen, B., Qin, P., & Zhuo, S. (2022). Smart contract generation assisted by AI-based word segmentation. *Applied Sciences*, 12(9), 4773.
- [19] Tupe, V., & Thube, S. (2025). AI Agentic workflows and Enterprise APIs: Adapting API architectures for the age of AI agents. *arXiv preprint arXiv:2502.17443*.
- [20] Xu, J., Li, J., Liu, Z., Suryanarayanan, N. A. V., Zhou, G., Guo, J., ... & Tei, K. (2024). Large language models synergize with automated machine learning. *arXiv preprint arXiv:2405.03727*.