AI Tools for Software Engineering

On this day in 1997, IBM's computer Deep Blue beat a world chess champion.

Abstract

Showing different references for AI integration into Software Development. 3 Artificial Intelligence tools are reviewed. Comparative results and conclusions are provided.
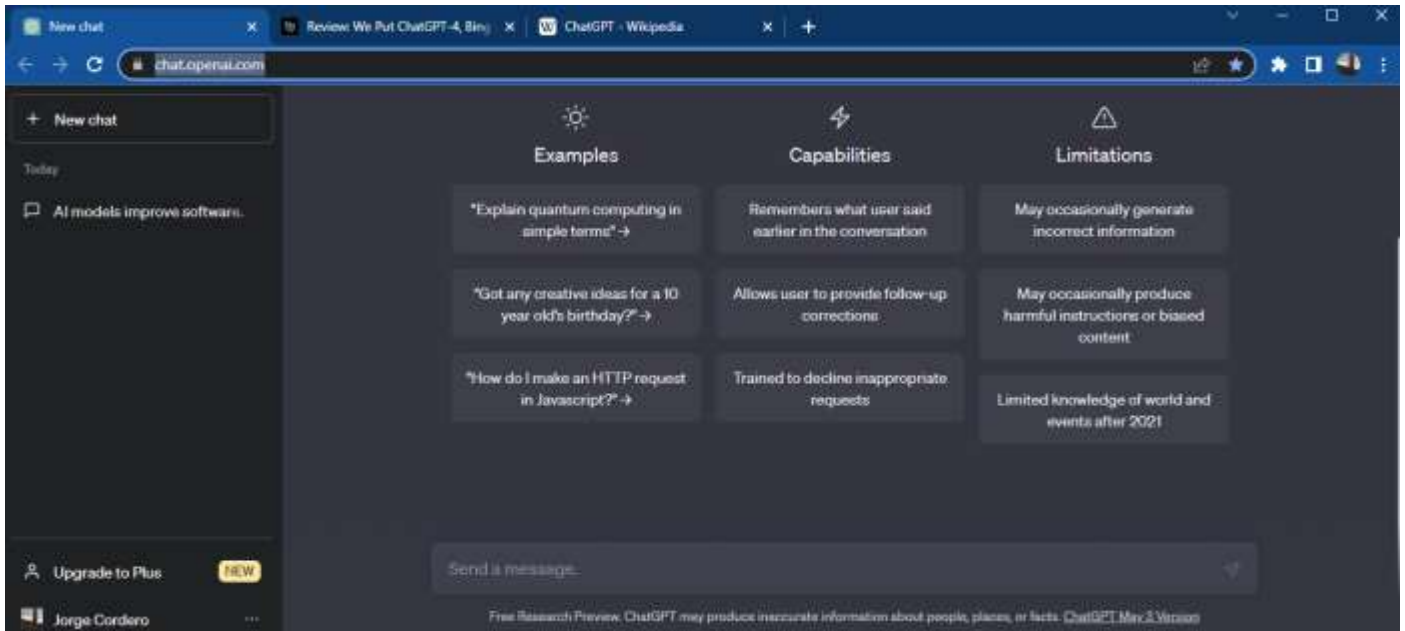
Jorge Pablo Cordero Hernández

May 13 2023

In this document I will present three different artificial intelligence (AI) tools applied to software development. The first 2 tools are "general purpose" for engineering and the third is a tool used to prototype machine learning applications. In each case, the characteristics, advantages and disadvantages of each technology are presented. Next, its characteristics are listed in a comparative table and finally a conclusion of the activity is presented.
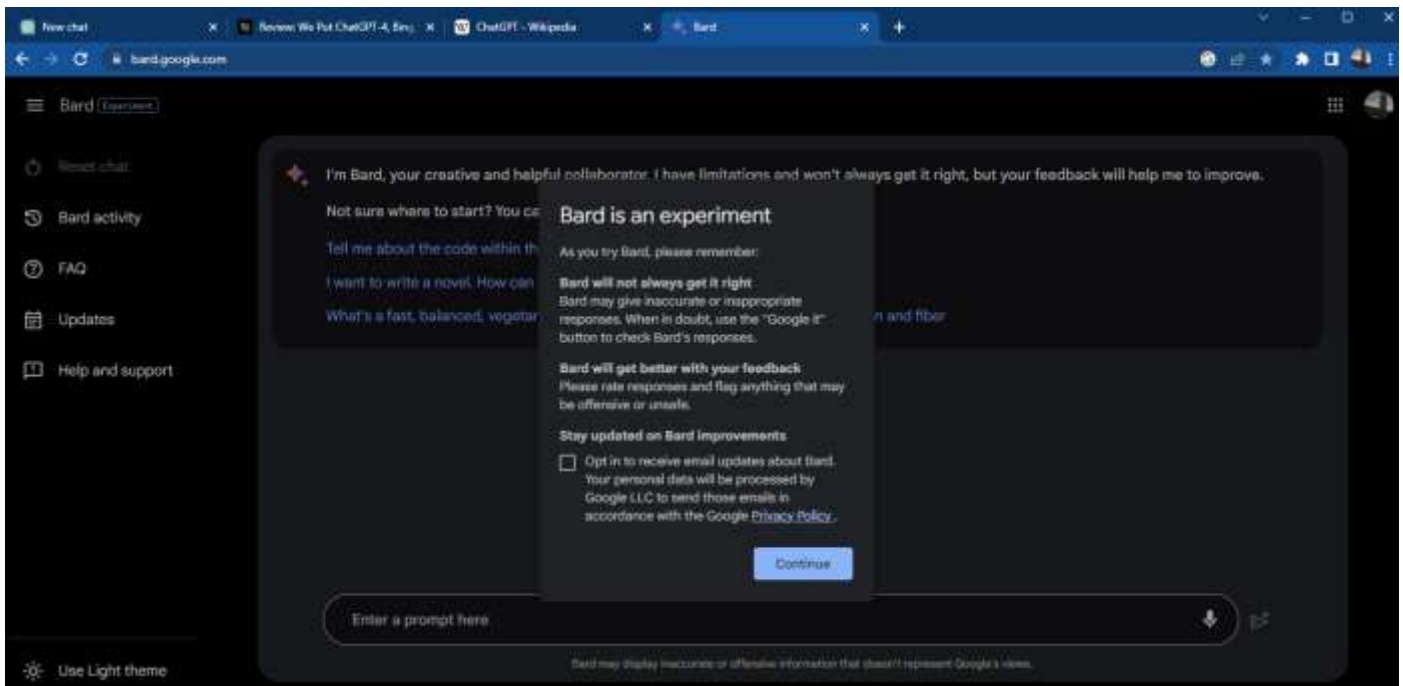
1. **ChatGPT [1]**

This program is probably the main reason why AI gained popularity again. It received too much publicity, its new functionality of being able to build code in any programming language from natural language (NLP) caused a lot of excitement. This software is a chatbot, the technology to make it work is based on machine learning models trained with Big Data. Although the entire nature of the algorithms used for its construction are not in the public domain, experts think that it could be Markov chain mechanisms and reinforcement learning along with neural networks.

One of the novelties of this chatbot is that it is now possible to use its "online learning" modality (in real time), to guide it in case it offers answers too far from the desired objective. Details on this setting can be found in [2]. The tool is available through a web interface on desktop computers or under a web service accessed by an "app". This application is so far the fastest growing in history [3], had more than one million users in the first five days after its launch.

There are different versions and licenses of this "app". Within software engineering there are plug-ins that can increase their specialization to solve "queries" in this context. However, openAI has shared its API for premium users. In this case, such a programmatic interface can be installed or communicated in the backend with custom developments by any individual or company.

Presenting enough test cases, this tool can complete code, suggest modules, to propose complete sequences of tasks in "devops" activities. Currently (since March 14, 2023) in version 4.0, 100 messages are available every 4 hours. Notably, the main limitation of supervised machine learning algorithms is their lack of self-demonstration. This family of algorithms serve as a black box, so they cannot explain or interpret the results that are produced.
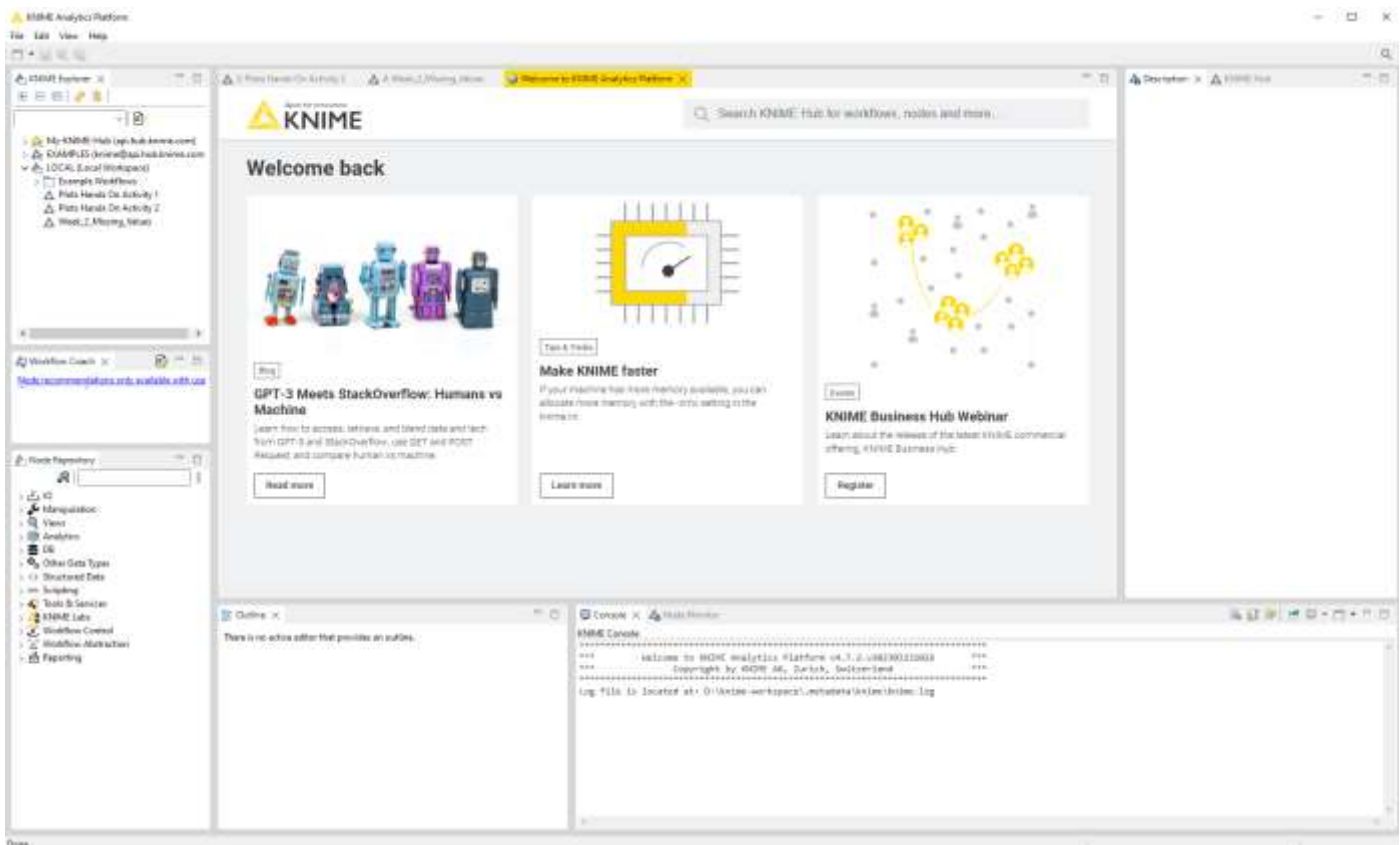
2. Bard [4]

Like chatGPT (from OpenAI), Bard is another chatbot developed by Google. The AI task they both perform is called generative artificial intelligence. Bard is now treated by Google as a running experiment and not a final product. This tool can also be used to generate code in any programming language. It should be said, that Google emphasizes the fact that this is an experiment and that Bard's programmatic results may be wrong at this time. So systems engineers or experts should evaluate and use their responses with caution.

It is interesting to acknowledge the fact that according to [5], the creators of Bard have programmed it to act committed to human values. This concept, its creator: Jacl Krawczyk calls it "alignment". There are improvements that will be offered soon, for example, a new interface that will be able to connect Bard with other Google and third-party applications (such as Adobe's Firefly suite).

The current version of Bard is distributed alongside the lightweight version of LaMDa (another AI framework for long language models). The main disadvantage of Bard is that it is not fully trained in many domains and as a result, it has flaws. There is no release process in place at this time, which all seems to indicate that Google pressured its team to bring an unfinished product to the public. This chatbot can also be used for the same tasks as chatGPT in the field of software engineering. A promising prospect is that Google is its developer, its enhancement announcements point to full interaction with its applications and cloud services. This will necessarily cause Bard to specialize and be a loyal competitor of chatGPT.
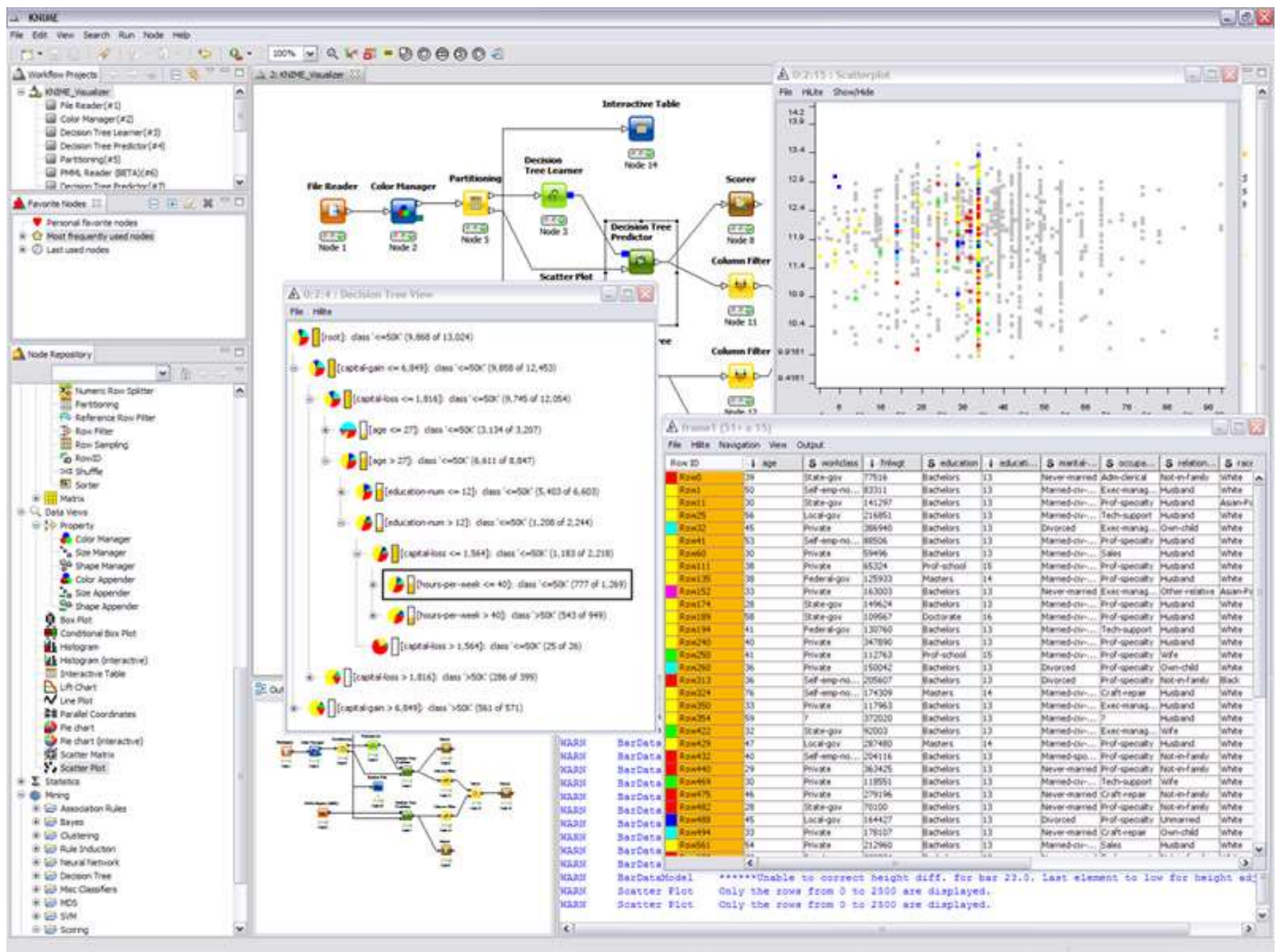
3. KNIME [7]

I think it's very important to keep in mind AI tools that aren't as popular as chatbots, but are the mainstay of building software based on machine learning tasks. KNIME is an open source tool for building machine learning models, both supervised and unsupervised. The idea is to simplify the task of modeling avoiding coding the algorithms in their entirety, the graphical interface allows the accelerated creation of prototypes, such as classifiers or linear regressors that can help pattern recognition. This is very important, since the development of AI can be generalized in any application in the future and the applications offered by large companies will tend to become more expensive depending on their use. For this reason, free software can help shape prototyping to integrate intelligent functionality into any piece of software.

This tool allows you to automate the entire life cycle ("pipelines") of a machine learning application. It has the flexibility to alter or customize any intermediate stage (e.g. preparation, extraction, transformation, loading, data normalization) at our convenience. This application is mature, and was initially used in pharmaceutical research. However, after several versions it has been extended to now be a multidisciplinary platform, among the areas where it has been used are: CRM customer management systems, marketing and customer data analysis, business intelligence, data and text mining. This tool has a graphical interface (as can be seen in the Figure shown below), and is the starting point for generating prototypes of AI software models that are then implemented in Java using libraries from other projects (for example, WEKA).

This tool allows you to automate the entire life cycle ("pipelines") of a machine learning application. It has the flexibility to alter or customize any intermediate stage (e.g. preparation, extraction, transformation, loading, data normalization) at our convenience. This application is mature, and was initially used in pharmaceutical research. However, after several versions it has been extended to now be a multidisciplinary platform, among the areas where it has been used are: CRM customer management systems, marketing and customer data analysis, business intelligence, data and text mining. This tool has a graphical interface (as can be seen in the Figure shown below), and is the starting point for generating prototypes of AI software models that are then implemented in Java using libraries from other projects (for example, WEKA).

KNIME provides the tools for loading data from files or databases through JDBC. It implements machine learning algorithms and can generate views with the results of AI applications.

There is a large number of work teams that use this tool as can be seen in [8, 9].

Comparison table: Features of different AI tools

Name	Company	Distribution	Versions	Characteristics	Operative System	Extensions	Advantages	Disadvantages
ChatGPT	OpenAI	Proprietary depending on the capabilities to be used: Basic level. Premium Service. APIs using REST.	GPT-3.5 GPT-3.5 Turbo AI GPT-4	NLP processing Autocompletion and code generation in any programming language. Musical composition Writing abstracts, essays and artistic works. Simulate text. Support for any automated operation. Emulation of operating systems. Resolution of any problem of scheduling and distribution of resources. Simulation of one or more agents/humans. Generator of simple computer games. Memorize the current conversation and use the interaction as semantic context.	Celulares: Android, Apple IOS, etc. PC: Through the client JavaGPT can be used in Windows 98, ME, 2000, XP, Vista, 7, 8, 10 and 11 [10] All operating systems: Through a browser or browser through REST requests on a Web page.	Whisper model APIs. Expedia. OpenTable. Zapier. Shopify. Slack. Wolfram. VSCode plug-in CodeGpt [11].	Interaction with the tool through natural message. It can be configured to be trained by means of text or "datasets". Accelerates the synthesis and elaboration of repetitive or trivial tasks. Generate code on any simple problem. For example, you can implement uncommon mathematical functions.	Prone to giving incorrect or meaningless answers if the wording of requests is ambiguous or training data is not available (called hallucination in long language models). Bias, the model tends to have errors that cannot be corrected by regularization. Both code and engineering applications are based on solutions to repetitive and sequential problems. Complex or "deep" problems are not properly modeled. You cannot comprehend or understand the results it generates.
		Owner. A single paid leave scheme.	1.0 LaMDA 2.0 PaLM 2	Currently all its functionalities are experimental.	All operating systems through a browser or	Plug in for Google Lens [12].	It uses other services such as Google Search to have a robust and	It may generate responses that contain inaccurate or misleading information. This despite

				Used for activity planning. Automatic writing of text and emails. Produces code functions and methods in any programming language. You will have the ability to generate images using the Adobe plug-in. It provides the same functionalities to be able to use characters and wildcards used in the lexicon and syntax to search in Google Search. It works freely with no waiting list for users, just have a Google account. Available in 180 countries.	browser through REST requests.	Adobe Firefly built-in image generator. Google Search with Bard. Services for Google Sheets, Docs and Maps. Extensions proposed in the future: Kayak, Spotify, OpenTable, ZipRecruiter, Instacart, Wolfram y Khan Academy. Plugins to support API calls to other programming languages such as: HTML/CSS, Java, C, C++, Dart, Kotlin etc.	experienced search for your training. Future integration with all services, cloud and software produced by Google. This will lead to the integration of Bard into all business/application areas of Google. Ease of providing more than one answer to questions. Unlike chatGPT, Bard can throw an array of answers for later comparison. It is proposed that it will be available in 26 languages shortly. Features to export chat sessions directly to Google Docs or Gmail. Options for customizing the appearance and colors of the page/interface.	presenting the results in a convincing way. You can generate biased answers based on the documents you take as training data. It cannot find a neutral criterion by consensus. Sometimes it can generate responses that seem to have personal opinions or emotions. Which is not the case. May give inappropriate or "tactless" responses in some circumstances. It is recommended for use only for people over 18 years of age.
Bard	Google						

KNIME	KNIME, Zurich	Individual: Free/Free. GNU General Public License. Owner: Community Hub, Basic, Standard, Enterprise.	Stable individual and proprietary: 4.7.2 / March 29, 2023. Individual stand alone. KNIME Business Hub.	Integral graphic solution to implement guided analytics, Functionalities to elaborate automatic reports. Business intelligence suite through a REST API for the paid version. Implementation of data mining algorithms. Implementation of Deep Learning algorithms with just one button. Data analysis in tables or graphs. Options for implementing text mining. Integration with databases and Big Data through JDBC libraries. Creation of data flows "pipelines" through implementations for atomic tasks in nodes. It has the ability to analyze data without preset limit. It is possible to do image mining with plug-ins.	Linux, macOS, Windows.	KNIME Active Learning [14, 15]. KNIME Amazon Athena Connector. KNIME Amazon Cloud Conectors. KNIME Amazon DynamoDB Nodes. KNIME Amazon Machine Learning Integration. KNIME Audio Processing. KNIME Azure Cloud Connectors. KNIME Base Chemistry Types and Nodes. KNIME Big Data Connectors. KNIME Conda Integration. KNIME Database. KNIME Databricks Integration.	Immediate development of machine learning solutions through a graphical interface. Reports are produced instantaneously. Automatic generation of graphs with professional presentation. It includes all AI algorithms that are popular or used in other tools like WEKA. Integration with programming languages such as Java, Python, R, etc. Offering the ability to integrate external code into the workflow. The visualization of the workflow is intuitive and can be explained to other stakeholders directly.	The interface is so easy to use that users with little knowledge of AI could confuse or generate erroneous reports. To build custom models based on code, the user needs to have specialized knowledge of machine learning and Java or Python. Which is a contradiction to the philosophy of its usability. Being a specialized tool, we don't have too many forums or user experience. The single version is open source; there is no ability to use KNIME functionalities or libraries through APIs to produce executable artifacts in other programming languages/environments. The only option to integrate KNIME into an external application is to use the paid version KNIME Business Hub.

				It provides tools for time series analysis. You can integrate Java code into your workflow. In addition, you can integrate other modules from products such as WEKA, H20.ai, Keras, Spark, R, Python, Ruby, LIBSVM, plotly, JFreeChart, ImageJ, Chemistry Development Kit, et …		KNIME Data Generation. KNIME Deep Learning4j Test Framework. KNIME Deep Learning – Keras Integration. Many others, please refer to [15] for more details.	Didactic, easy to understand as a tool to learn machine learning. Fast and optimized. Enables rapid prototyping of machine learning software.	With this we can make http requests using a REST API on the KNIME server. This represents a monetary, temporary operational cost, and the details of our application are exposed to possible vulnerabilities in the network.

Conclusion

It is difficult to make a complete and accurate value judgment of what were the reasons why AI gained popularity again during the year 2022. Possibly it was the large amount of advertising invested in openAI's chatGPT. I mention this because both the machine learning techniques used and the application of natural language processing have been available since the nineties. Even the algorithms are the same, the only difference is that before we did not have massive hardware in the cloud for data storage and processing. Nor was there the economic initiative to invest and risk capital in applications that could compete with other products of technological giants.

It's no surprise that autonomous NLP processors or chatbots can now offer to fix, autocomplete, or provide finished implementations of code. It is enough to list the different frameworks in Java such as Lombok, Spring, Swagger, etc. (or the extreme ease that Python provides to implement from web applications to the training and execution of complex AI algorithms with only two lines of code) to realize that, the tendency is to make programmers could program software solutions in a simpler and therefore faster way. This produces agile or rapid developments, but also leads to a lack of understanding of the tasks that are being implemented.

Logically, when code can be factored and its details are delegated to libraries or APIs in their entirety, then any classification algorithm can systematically learn the relationships between requirements and outcomes. This kind of problem is what these algorithms have solved for decades. Thus producing complete implementations in the form of programmatic tags grouped in code. The information related to these technologies is extremely easy to obtain, from any search engine. For technical details on its implementation, just go to the AI bibliographies available for more than 30 years.

The research helped me get in touch with the trendy chatbots: chatGPT and Bard. It was useful to be able to have a vision of the scope of these tools. These can be used to aid in the task of software development. It is enough to express yourself in natural language to get more direct answers than we can find under a Google Search search. KNIME is one of many AI tools I used before, as data analysis is a fundamental task in modern applications.

References

[1] Chat GPT. Open AI. https://chat.openai.com/, https://openai.com/blog/chatgpt. Read on May the 14th, 2023.

[2] Mollick, Ethan. CHATgpt Is a Tipping Point for AI. Harvard Business Review. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai. Read on May the 14th, 2023.

[3] Ortiz, Sabrina. What is ChatGPT and why does it matter? Here's what you need to now. ZDNET. April 18, 2023. https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-everything-you-need-to-know/. Read on May the 14th, 2023.

[4] Google Bard. https://bard.google.com/ .Read on May the 14th, 2023.

[5] Shankland, Stephen. Google's Bard Chatbot Opens to the Public. CNET. May 10, 2023. https://www.cnet.com/tech/computing/googles-bard-chatbot-opens-to-the-public/. Read on May the 14th, 2023.

[6] Pichai, Sundar. Google Blog. February 2023. https://blog.google/technology/ai/bard-google-ai-search-updates/. Read on May the 14th, 2023.

[7] KNIME. The Konstanz Information Miner. Open for Innovation: The case of Openness. https://www.knime.com/about. Read on May the 14th, 2023.

[8] Knime Analytics Platform Reviews. Software Advice. https://www.softwareadvice.com/bi/knime-analytics-platform-profile/reviews/. Read on May the 14th, 2023.

[9] Knime Analytics Platform Ratings Overview. Gartner Peer Insights. https://www.gartner.com/reviews/market/dsml-engineering-platforms/vendor/knime/product/knime-analytics-platform. Read on May the 14th, 2023.

[10] Endicott, Sean. JavaGPT brings chatGPT to support ancient versions on Windows. https://www.windowscentral.com/software-apps/javagpt-brings-chatgpt-to-support-ancient-versions-of-windows. Read on May the 14th, 2023.

[11] CodeGPT: The VSCode Extension with ChatGPT-Like Functionalities. https://medium.com/geekculture/codegpt-the-vscode-extension-with-chatgpt-like-functionalities-783323a916c3. Read on May the 14th, 2023.

[12] Sharma, Upanishad. 10+ Best New and Upcoming Google Bard Features. https://beebom.com/google-bard-ai-best-features/. Read on May the 14th, 2023.

[13] Google lists five 'limitations' of its chatGPT rival Bard. https://timesofindia.indiatimes.com/gadgets-news/google-lists-5-known-limitations-of-llm-based-ai-chatbots-like-bard/articleshow/98943051.cms. Read on May the 14th, 2023.

[14] KNIME forum. https://forum.knime.com/about. Read on May the 14th, 2023.

[15] KNIME Extensions. https://hub.knime.com/knime/extensions. Read on May 14th, 2023.