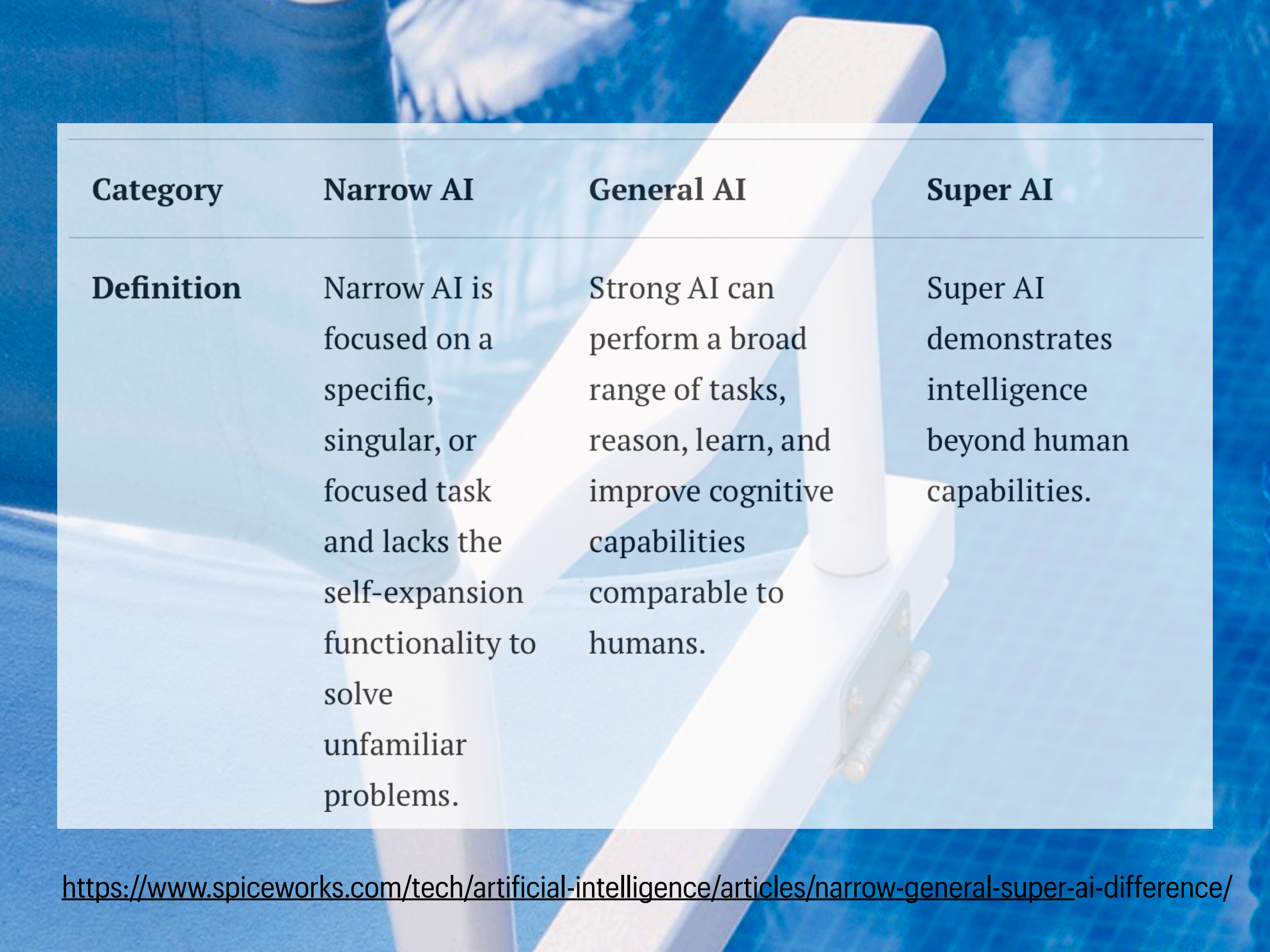


**Artificial General Intelligence
obsoletes**

**software reverse/re-engineering research
as we know it!?**

WSRE 2023 — PANEL

RALF LÄMMEL, UNIVERSITÄT KOBLENZ



Category	Narrow AI	General AI	Super AI
Definition	Narrow AI is focused on a specific, singular, or focused task and lacks the self-expansion functionality to solve unfamiliar problems.	Strong AI can perform a broad range of tasks, reason, learn, and improve cognitive capabilities comparable to humans.	Super AI demonstrates intelligence beyond human capabilities.

Language models are nearly AGIs but we don't notice it because we keep shifting the bar

by philosophybear 8 min read 30th Dec 2022 13 comments

<https://www.lesswrong.com/posts/myCA3AzopBhnYB6P/language-models-are-nearly-agis-but-we-don-t-notice-it>

^
94
v

Does the Textual Corpus for Large Language Models Have Enough Information to Train an AGI?

Given the size and contents of our textual data, probably.

<https://shakoist.substack.com/p/does-the-textual-corpus-for-large>

**Language models
support
software reverse/re-engineering research!**

WSRE 2023 — PANEL

RALF LÄMMEL, UNIVERSITÄT KOBLENZ

Additional resources:

- [Supporting Blog Post](#)
- [Supporting Annotated Chats with ChatGPT](#)

What are the principal types of research?

According to ChatGPT ...

- ***Descriptive*** — to describe a particular phenomenon or situation.
- ***Exploratory*** — to gain a general understanding of a topic or problem.
- ***Correlational*** — to examine the relationship between 2+ variables.
- ***Experimental*** — to manipulate variables & to observe effect on others.
- ***Qualitative*** — to understand experiences & perspectives of individuals.
- ***Quantitative*** — to collect & analyze data to test hypotheses.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Descriptive research can be used in software reverse engineering to describe the **characteristics of a particular software system or codebase**, such as the size, complexity, and structure of the code.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Exploratory research can be used in software reverse engineering to gain a ***general understanding of a software system or codebase***, and to identify key variables and relationships that can be explored further.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Correlational research can be used in software reverse engineering to examine the relationship between different software metrics, such as **code complexity and maintainability**.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Experimental research can be used in software re-engineering to manipulate one or more variables, such as *code refactoring techniques*, and to observe the effect on *software quality metrics*.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Qualitative research can be used in software reverse engineering to *understand the experiences and perspectives of software developers and maintainers*, and to gain insights into the challenges and opportunities involved in software maintenance and evolution.

Types of research in SWRe/ReverseE

According to ChatGPT ...

- *Descriptive* — to describe a particular phenomenon or situation.
- *Exploratory* — to gain a general understanding of a topic or problem.
- *Correlational* — to examine the relationship between 2+ variables.
- *Experimental* — to manipulate variables & to observe effect on others.
- *Qualitative* — to understand experiences & perspectives of individuals.
- *Quantitative* — to collect & analyze data to test hypotheses.

Quantitative research can be used in software re-engineering to *collect and analyze numerical data, such as software metrics and performance measurements, and to test hypotheses about the effectiveness of different re-engineering approaches.*

ChatGPT goes SW reverse/re-engineering

What you always wanted to know ...

- **How can the AI support reverse/re-engineering?**
- **Can the AI “explain” methodology options?**
- **Can the AI cite papers on topics?**
- **Can the AI sample/process repositories?**
- **Is the AI self-aware when it comes to its research abilities?**

Exploration of ChatGPT's capabilities in chats

We designed chats as follows:

1. Let the AI propose a **methodology** for a relatively specific research scenario in the context of **software reverse engineering**: some sort of **architecture recovery for callbacks** to be evaluated with a methodology as in Mining Software Repository (MSR). Advance the communication to determine the AI's ability to support automation of certain steps of the methodology.
2. Let the AI propose a structured **implementation** (approach) of a relatively specific scenario in the context of **software re-engineering**: some sort of **language migration** (Cobol to Java). Advance the communication to determine the AI's ability to generate some central parts of the implementation meeting a technology (ANTLR) and design choice (grammar-to-grammar mapping).
3. Let the AI propose a list of **research challenges in software re- and reverse engineering** while also demanding eventual specialization of a selected challenge (here: **code deobfuscation**) down to the level of code-based illustration and references to suitable literature.
4. Let the AI reflect on an **AI's capabilities to carry out more or less autonomously research** on software re- and reverse engineering. We want to see what **limitations** the AI identifies for itself and how it assesses the need for **cooperation with human counterparts**.

ChatGPT's merits in SW reverse/re-engineering

Our findings based on chatting with ChatGPT

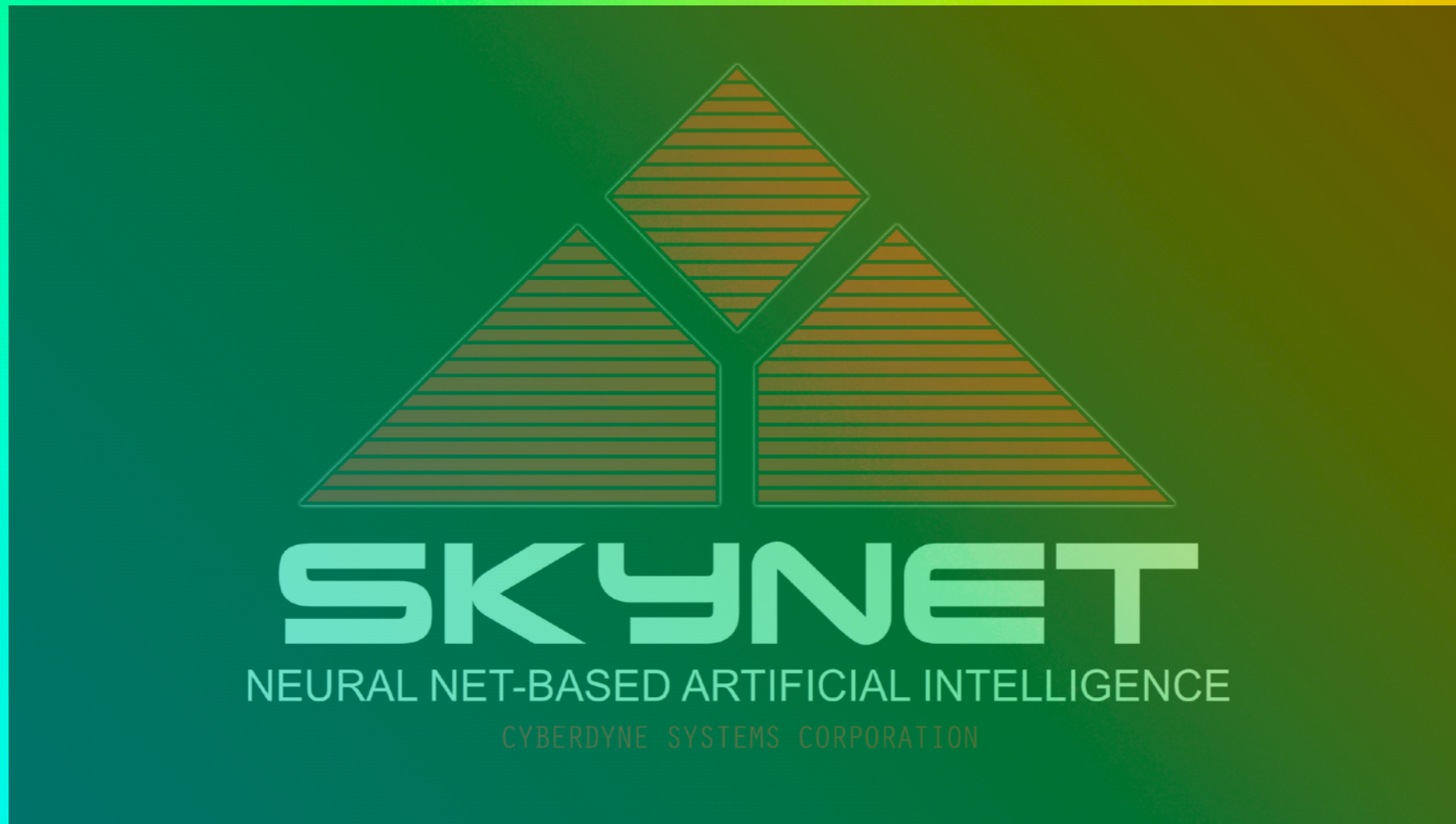
Capabilities

- **Generate custom methodology based on provided constraints**
- **Generate non-trivial code (e.g., for language migration)**
- **Generate research challenges with increasing specificity and illustrations**
- **Generate AI abilities that would help automating SWReverse/ReE**

Limitations

- **Cannot custom filter GitHub repos**
- **Cannot align NL and code structures reliably enough**
- **Cannot perform basic citation**
- **Pushes "cooperation AI + Human" in unconvincing manner**

Pessimistic summary



https://upload.wikimedia.org/wikipedia/commons/4/4d/Skynet_Terminator_logo.png

Focused, stubborn summary



Julia Turc

Jul 13, 2022 · 7 min read · ✦ Member-only · 🎧 Listen



Language Models that Act, Not Just Talk

How Google's Minerva promises a future where machines can act

Large Language Models like GPT-3 have mostly been used for the same task they were trained to perform: text generation. However, language is merely a means to an end. In the coming years, we will see an inflection where models **act**, not just **talk**.

Realistic summary



Somnath Singh

Jan 20 · 8 min read · Member-only · Listen



Coding Won't Exist In 5 Years. This Is Why

Those who won't adapt would cease to exist.



[Source](#)

(More) pointers

- My blog post on the subject.
- More on language models:
 - <https://paperswithcode.com/method/codet5>
 - <https://github.com/microsoft/CodeBERT>
 - <https://www.techtarget.com/whatis/feature/Bard-vs-ChatGPT-Whats-the-difference>
 - <https://ai.googleblog.com/2022/06/minerva-solving-quantitative-reasoning.html>
 - <https://www.digitaltrends.com/computing/chatgpt-4-everything-we-know-so-far/>
 - <https://towardsdatascience.com/beyond-text-generation-language-models-that-act-not-just-talk-127236c0976d>
- One more optimistic outlook:
 - <https://javascript.plainenglish.io/coding-wont-exist-in-5-years-this-is-why-6da748ba676c>

ChatGPT's merits in SW reverse/re-engineering

ChatGTP pushes "cooperation AI + Human" in unconvincing manne

As a language model and AI, I can collaborate with a computer science student on their research project on software reverse and re-engineering in various ways, such as providing information and explanations related to software engineering concepts, analyzing software code and data, and generating scripts and programs to automate certain tasks.

However, there may be certain steps of attempted automation that would fail if the student tried to delegate work to me. Some examples of such steps include:

- Lack of domain-specific knowledge: ...**
- Limited access to data and computational resources: ...**
- Limited ability to validate and test generated scripts and programs: ...**
- Limited ability to adapt to new and evolving software engineering concepts: ...**

Therefore, while I can collaborate with a computer science student on their research project on software reverse and re-engineering in various ways, there may be certain steps of attempted automation that would fail.