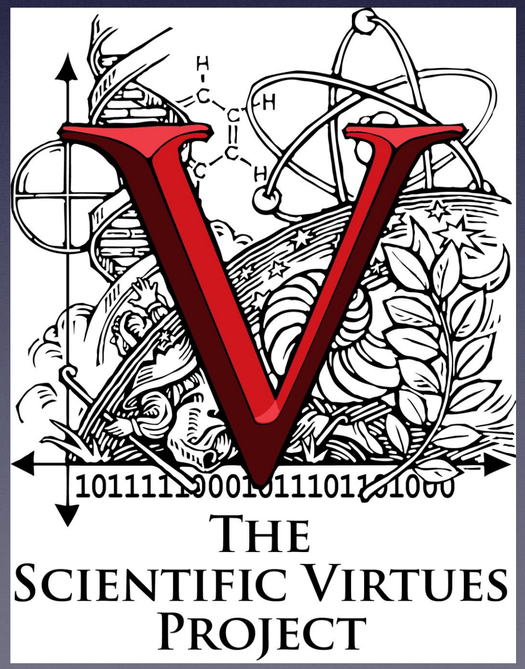
AI Chatbots & the Ethics of Authorship



Robert T. Pennock
Lyman Briggs College
Computer Science & Engineering
Philosophy
Ecology, Evolution & Behavior
Michigan State University

WCRI
6/3/2024
Athens, Greece

From Lovelace to LLMs

Is AI alive?
Is AI sentient?
Is AI intelligent?
Is AI an author?

Magic Studio

Professional organizations call for ethical standards of conduct

- National Academy of Sciences, Institute of Medicine, and National Academy of Engineering issue statement in 1994 decrying lack of progress in instituting ethical standards for scientific research.
- Cite “inappropriately assigning authorship to research papers” as one sort of misconduct.
- Today’s question: Should AI bot be listed as an author on a scientific research paper?

Authorship

Caveats regarding application of the concept in science

- (1) Writing of a research report is the only aspect of the scientist's job to which idea of authorship applies, and this is the last and the least scientific part of the work.
- (2) "Author" more than not carries the connotation of "creator," which does not fit the scientific context, despite claims to the contrary by postmodernist critics.
- (3) Linked etymologically to the idea of authority and thereby to an unscientific, legislative, notion of justification.

Ethical Misconduct? or Poor Credit Attribution Model?

- THESIS: While cases of intentional misconduct do occur, many of the moral problems are unintentional. Rather they are the result of the simplistic and archaic authorship model of credit attribution itself and could be solved by a more explicit and precise model.

- Pennock, R.T. “Inappropriate Authorship in Collaborative Scientific Research” *Public Affairs Quarterly*, Vol. 10, Number 4, pp. 379-393, October 1996.

Proposed Solution
“The Credits” Section

- **Rather than the current list of “authors,” papers that report collaborative work should have a credits section that names the collaborators according to the roles they played or the contributions they made to the research.**

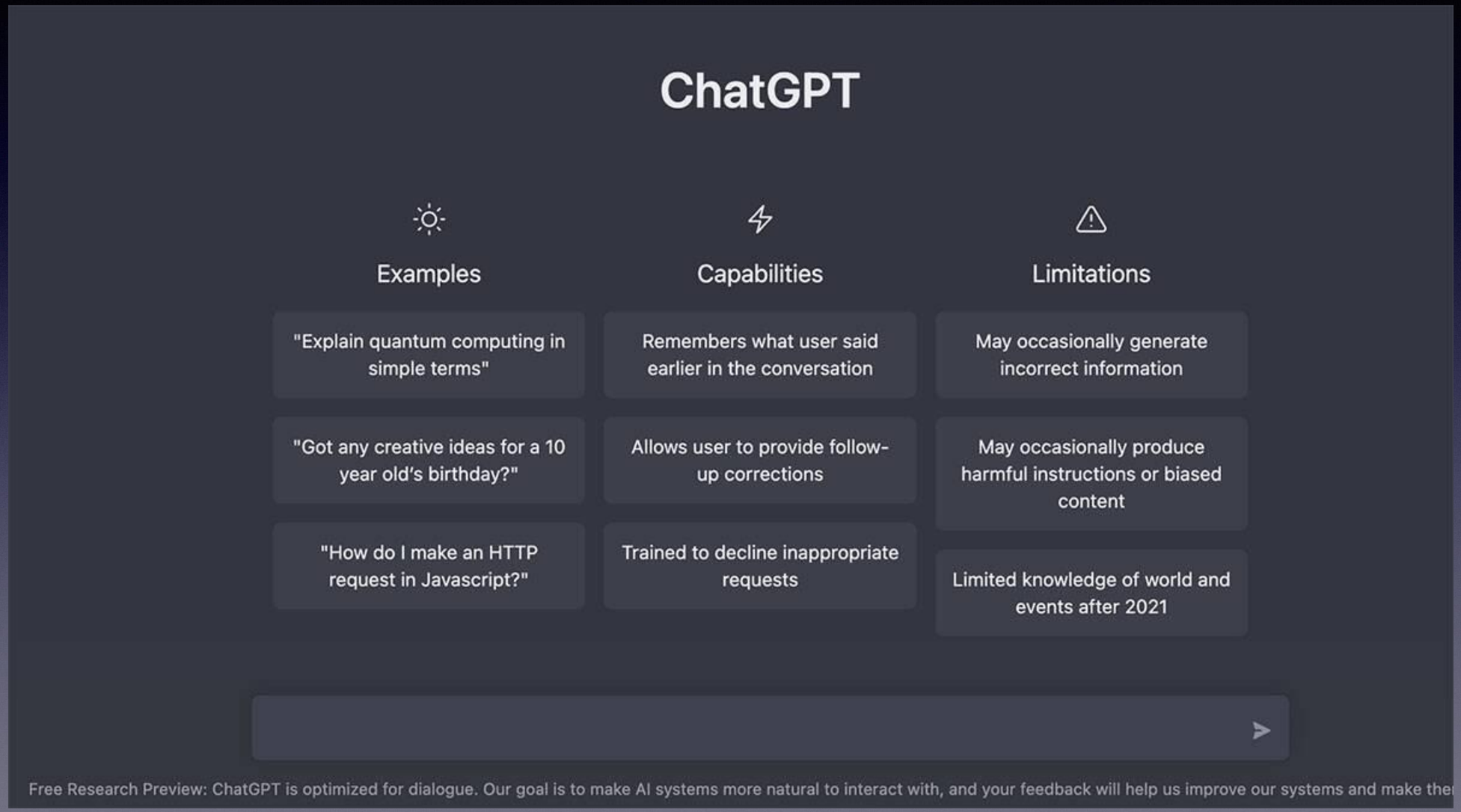
- Some roles will be common across disciplines, such as “Principle Investigator” (the one or a few researchers who oversee and take responsibility for the entire project)

- Others may be specific to the study, such as “Laser Spectrographer,” “Statistician” or “Virus samples contributed by….”

How this would help
- An explicit credits section…

- Avoids the misleading connotations of “author”
- Provides information that is in keeping with and conducive to truth-seeking.
- Allows just distribution of credit
- Allows just distribution of blame
- Allows researchers to put their signature to those aspects of the research for which they take responsibility

How do LLMs work?



LLM Ethical Concerns

- Copyright violation / plagiarism in training set data
- New mode of cheating
- Halucinations / confabulations
- Biases

Screenshot: u/s-p-o-o-k-i—m-e-m-e / Reddit / Microsoft

The Authorship Credit

Praise and Blame

Responsibility

- Difference between being the causally responsible for E and being morally responsible for E
- Fact vs value
- Being responsible as a moral concept involves taking on the onus of answering—responding—if questioned

Cf. taking on the burden of justification

Authorship in Science as Analogous to Endorsing a Check

- A scientific paper is not an act of creation but a report of evidence for a discovery.
- Signing off on a report indicates that I have performed the requisite tests and will stand by them.
- It is a taking on of moral responsibility., which AI bots can’t (yet) do.
- AI bots should not be listed as authors, but as tools.

PUBLIC AFFAIRS QUARTERLY
Volume 10, Number 4, October 1996

INAPPROPRIATE AUTHORSHIP IN COLLABORATIVE SCIENCE RESEARCH

Robert T. Pennock

A recent statement from the National Academy of Sciences, the Institute of Medicine and the National Academy of Engineering decried the lack of progress in instituting ethical standards for scientific research, and among the sorts of misconduct they mentioned was "inappropriately assigning authorship to research papers" (Hilts 1994).

Kristin Shrader-Frechette terms this unethical practice "loose authorship" in her *Ethics of Scientific Research* and defines it as "inserting or removing names of pers
work", but she does not
sial" form of deception
paper I take up this iss
scientific research and e
current situation in whic
and analyzes specific for
three major ethical pri
tice—that are violated ir
demand for proper attrib
tions and show how the
searchers to satisfy the
attribution strategies tha
defend them against pos
the conclusions and ma
facilitate change. My c
ship model itself, with
blame for many of the e
ties could be avoided i
journals adopted an exp
bution conventions.

Science and Engineering Values

AI and Responsible Authorship

Why my chatbot is not (yet) a coauthor.

Robert T. Pennock

Suppose I do the experiments but use an artificial intelligence chatbot to write the report; should I list it as an author? If I only use the chatbot to flag typos or suggest fixes for grammatical errors, that question would never arise. But what if, to save time, I have the AI write the literature review section summarizing a set of articles I gave it? Now the words on the page are not my own. More significantly, what if I give it my experimental data to analyze and write up? As AI increases in power and capabilities, does it deserve credit as a coauthor?

From Lovelace to LLMs

In 1843, Ada Lovelace published what was arguably the first computer program, showing how an analytical engine—as mathematician Charles Babbage called his yet-unbuilt digital mechanism—could calculate a common sequence of rational numbers called Bernoulli numbers. A computer program is just a step-by-step procedure, but Lovelace's algorithm could do something that at the time only a person could. An algorithm may run on a mechanical device with gears, an electrical device with circuits, or on an abstract writing instrument and roll of paper that moves based on symbols written on it—a Turing machine, named after computer pioneer Alan Turing. The idea of artificial intelligence is that such artifacts can in principle be able to exhibit recognizable, if perhaps not exactly human, intelligent activity.

As a PhD student in the late 1980s, I worked with Herbert Simon, the Nobel Prize–winning polymath known as the father of AI for his pioneering theoretical and empirical work that founded the field. Simon argued that AI should be analyzed in terms of symbolic reasoning. I also heard computer scientist and cognitive psychologist Geoffrey Hinton, now called the godfather of AI, argue for and demonstrate early results of an alternative "connectionist" approach that focused instead on statistical associations in *artificial neural networks* (ANNs). ANNs were modeled on brain structures, with varying weights of connections between nodes governing the processing from input to output.

The write-up serves a vital function because it reports the evidence, but authoring is not the core part of research.

The relative merits of these approaches made for vibrant debate and drove interesting research. For example, symbolic AI researchers analyzed the rules of expert reasoning and devised programs to simulate them. Hinton, who later received the A. M. Turing Award, and other connectionists devised better ways to train ANNs. For years, practical applications in both symbolic and connectionist AI always seemed beyond the horizon, but the early 2020s saw rapid advances in their capabilities.

By training an attention-based transformer model on massive amounts of text gathered from the internet, the weights of an ANN can be adjusted so that it becomes an adroit text manipulator. Such *large language models* (LLMs) take a text prompt as an input and then generate a string of output text based on predictions of what words should follow what came before. LLMs generally lack symbolic structure and are not yet very good at math, but various hybrid systems attempt to combine the strengths of both symbolic and connectionist models. Today LLMs can generate a convincing essay about Bernoulli's mathematical discovery. Lovelace would be impressed.

The term *computer* originally referred to a person who performed mathematical calculations—computations. When
comp
clerica
term.
autho
enoug
an aut

Authc
To ana
entific
sider t
first p
if only
molog
thereb
notion

QUICK TAKE

The use of artificial intelligence in the development of research papers raises the ethical question of whether AI tools should receive coauthor credit.

Responsible scientific authorship is less about authoring the text of a research paper than it is authorizing the paper as a fair representation of the evidence supporting its findings.

AI agent
reference
they are
the resea

Magic Studio