

Introduction



2023 Expert Survey on Progress in AI

Welcome. We are conducting a study of progress in artificial intelligence and are interested in your understanding of developments in the field.

Our estimated median time for completing this survey is around 16 minutes. We will publish summary results and a survey dataset, but in them your responses will be kept **anonymous**.

Many of the questions involve substantial uncertainties. Please just give us your **current best guesses**.

There are no known risks associated with this study. Although this study may not benefit you personally, we hope that our results will add to the knowledge about progress in AI technology.

Katja Grace will control your data, which will be transferred to the United States. It will also be accessed by

colleagues and contractors involved in processing the data, or with data storage or processing services. It could be shared with public authorities, if we are lawfully ordered to do so. If you have questions about your rights as a research participant, you may contact Katja Grace: 412-304-6258 katja@aiimpacts.org.

We will publish summary results and release an anonymized dataset. Your data will be anonymous in all public publications. We will use your personal data primarily for the purposes of this research project, but may conduct related followup projects which also use the personal data collected for this project.

If you take this survey, in addition to the information you will directly submit to us, we will also receive data reported automatically by the survey software about your location, IP address, and survey timing.

Participation in this study is completely voluntary. You are free to decline to participate and to end participation at any time for any reason.

By continuing to the next page, you agree to participate in the survey.

Note: Due to technical limitations and randomization, **once you navigate forward from a page, you will typically be unable to return to it.** Should any issues arise, please reach out to harlan@aiimpacts.org. This contact information will be provided again at the end of the survey.

HLMI basic

1 of 7

The following questions ask about **'high-level machine intelligence'** (HLMI).

Say we have **'high-level machine intelligence'** when unaided machines can accomplish every task better and more cheaply than human workers. Ignore aspects of tasks for which being a human is

intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*

For the purposes of this question, assume that human scientific activity continues without major negative disruption.

How many years until you expect:

- a 10% probability of HLMI existing? years
- a 50% probability of HLMI existing? years
- a 90% probability of HLMI existing? years

For the purposes of this question, assume that human scientific activity continues without major negative disruption.

How likely is it that HLMI exists:

- in 10 years? %
- in 20 years? %
- in 40 years? %

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

HLMI via jobs - fixed probabilities

Say an occupation becomes **fully automatable** when unaided machines can accomplish it better and more cheaply than human workers. Ignore aspects of occupations for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*

We want to know how many years you think will pass before the following present-day occupations will be **fully automatable**. Please tell us your best guess of when you think there will be a small chance (10% chance), a roughly even chance (50% chance), and a high chance (90% chance).

	Years until small chance (10%)	Years until even chance (50%)	Years until high chance (90%)
Truck driver	<input type="text"/>	<input type="text"/>	<input type="text"/>
Surgeon	<input type="text"/>	<input type="text"/>	<input type="text"/>
Retail salesperson	<input type="text"/>	<input type="text"/>	<input type="text"/>
AI researcher	<input type="text"/>	<input type="text"/>	<input type="text"/>

What is an existing human occupation that you think will be among the final ones to be **fully automatable**?

Remember to consider feasibility, not adoption.

How many years do you expect to pass before you think there is a small/even/high chance that this occupation will be **fully automatable**?

Small chance (10%) years

Even chance (50%) years

High chance (90%) years

Say we have reached '**full automation of labor**' when all occupations are fully automatable. That is, when for any occupation, machines could be built to carry out the task better and more cheaply than human workers.

In how many years do you expect **full automation of labor**, with small/even/high chance?

Small chance (10%) years

Even chance (50%) years

High chance (90%) years

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

HLMI via jobs - fixed years

Say an occupation becomes **fully automatable** when unaided machines can accomplish it better and more cheaply than human workers. Ignore aspects of occupations for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*

We want to know how likely you think it is that the following present-day occupations will be **fully automatable** at future dates. Please tell us your best guess of the chance that they will be **fully automatable** within the next 10 years, within the next 20 years, and within the next 50 years.

	% chance in 10 years	% chance in 20 years	% chance in 50 years
Truck driver	<input type="text"/>	<input type="text"/>	<input type="text"/>
Surgeon	<input type="text"/>	<input type="text"/>	<input type="text"/>
Retail salesperson	<input type="text"/>	<input type="text"/>	<input type="text"/>
AI researcher	<input type="text"/>	<input type="text"/>	<input type="text"/>

What is an existing human occupation that you think will be among the final ones to be **fully automatable**?

Remember to consider feasibility, not adoption.

How likely do you think it is that this occupation will be **fully automatable** within the next 10/20/50 years?

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Say we have reached ‘**full automation of labor**’ when all occupations are **fully automatable**. That is, when for any occupation, machines could be built to carry out the task better and more cheaply than human workers.

How likely do you think it is that full automation of labor will happen within the next 10/20/50 years?

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Intelligence explosion

2 of 7

The following questions ask about '**high-level machine intelligence**' (HLMI).

Say we have '**high-level machine intelligence**' when unaided machines can accomplish every task better and more cheaply than human workers. Ignore aspects of tasks for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*

Assume that HLMI will exist at some point.

How likely do you then think it is that the rate of global technological improvement will dramatically increase (e.g. by a factor of ten) as a result of machine intelligence:

Within **two years** of that point? % chance

Within **thirty years** of that point? % chance

Assume that HLMI will exist at some point.

How likely do you think it is that there will be machine intelligence that is **vastly better** than humans at all professions (i.e. that is vastly more capable or vastly cheaper):

Within **two years** of that point? % chance

Within **thirty years** of that point? % chance

Some people have argued the following:

If AI systems do nearly all research and development, improvements in AI will accelerate the pace of technological progress, including further progress in AI.

Over a short period (less than 5 years), this feedback loop could cause technological progress to become more than an order of magnitude faster.

How likely do you find this argument to be broadly correct?

Quite unlikely
(0-20%)



Unlikely
(21-40%)



About even
chance
(41-60%)



Likely
(61-80%)



Quite likely
(81-100%)



Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Explainability 2

For typical state-of-the-art AI systems in 2028, do you think it will be possible for users to know the true reasons for systems making a particular choice? By “true reasons” we mean the AI correctly explains its internal decision-making process in a way humans can understand. By “true reasons” we do **not** mean the decision itself is correct.

Very unlikely
(<10%)



Unlikely
(10-40%)



Even odds
(40-60%)



Likely
(60-90%)



Very likely
(>90%)



Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Scenarios

How would you rate the level of concern these scenarios deserve over the next thirty years?

	No concern	A little concern	Substantial concern	Extreer conce
A powerful AI system has its goals not set right, causing a catastrophe (e.g. it develops and uses powerful weapons)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
AI makes it easy to spread false information, e.g. deepfakes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
AI systems manipulate large-scale public opinion trends	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
AI systems worsen economic inequality by disproportionately benefiting certain individuals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
AI systems with the wrong goals become very powerful and reduce the role of humans in making decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
AI lets dangerous groups make powerful tools (e.g. engineered viruses)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	No concern	A little concern	Substantial concern	Extreer conce
Bias in AI systems makes unjust situations worse, e.g. AI systems learn to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

AI systems learn to discriminate by gender or race in hiring processes

Near-full automation of labor makes people struggle to find meaning in their lives.

Near-full automation of labor leaves most people economically powerless

People interact with other humans less because they are spending more time interacting with AI systems

Authoritarian rulers use AI to control their population

Other

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Value

3 of 7

The following questions ask about '**high-level machine intelligence**' (HLMI).

Say we have '**high-level machine intelligence**' when unaided machines can accomplish every task better and more cheaply than human workers. Ignore aspects of tasks for which being a human is intrinsically advantageous, e.g. being accepted as a jury member. *Think feasibility, not adoption.*

Assume for the purpose of this question that HLMI will at some point exist. How positive or negative do you expect the overall impact of this to be on humanity, in the long run?

Please answer by saying how probable you find the following kinds of impact, with probabilities adding to 100%:

Extremely good (e.g. rapid growth in human flourishing)	<input type="text" value="0"/>	%
On balance good	<input type="text" value="0"/>	%
More or less neutral	<input type="text" value="0"/>	%
On balance bad	<input type="text" value="0"/>	%
Extremely bad (e.g. human extinction)	<input type="text" value="0"/>	%
Total	<input type="text" value="0"/>	%

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Causes of AI Progress

4 of 7

The next questions ask about the sensitivity of progress in AI capabilities to changes in inputs.

'Progress in AI capabilities' is an imprecise concept, so we are asking about progress as you naturally conceive of it, and looking for approximate answers.

Imagine that over the past decade, only **half as much researcher effort** had gone into AI research. For instance, if there were actually 1,000 researchers, imagine that there had been only 500 researchers (of the same quality).

How much less progress in AI capabilities would you expect to have seen?

e.g. If you think progress is linear in the number of researchers, so 50% less progress would have been made, write '50'. If you think only 20% less progress would have been made write '20'.

% less

Over the last n years the cost of computing hardware has fallen by a factor of 20. Imagine instead that **the cost of computing hardware had fallen by only a factor of 5** over that time (around half as far on a log scale).

How much less progress in AI capabilities would you expect to have seen?

e.g. If you think progress is linear in $1/\text{cost}$, so that $1 - 5/20 = 75\%$ less progress would have been made, write '75'. If you think only 20% less progress would have been made write '20'.

% less

Imagine that over the past decade, there had only been **half as much effort put into increasing the size and availability of training datasets**. For instance, perhaps there are only half as many datasets, or perhaps existing datasets are substantially smaller or lower quality.

How much less progress in AI capabilities would you expect to have seen?

e.g. If you think 20% less progress would have been made, write '20'

% less

Imagine that over the past decade, AI research had **half as much funding** (in both academic and industry labs). For instance, if the average lab had a budget of \$20 million each year, suppose their budget had only been \$10 million each year.

How much less progress in AI capabilities would you expect to have seen?

e.g. If you think 20% less progress would have been made, write '20'

% less

Imagine that over the past decade, there had been **half as much progress in AI algorithms**. You might imagine this as conceptual insights being half as frequent.

How much less progress in AI capabilities would you expect to have seen?

e.g. If you think 20% less progress would have been made, write '20'

% less

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these question? (optional)

HLMI Outside view

4 of 7

Which AI research area have you worked in for the longest time?

How long have you worked in this area?

 years

Consider three levels of progress or advancement in this area:

A. Where the area was when you started working in it

B. Where it is now

C. Where it would need to be for AI software to have roughly human level abilities at the tasks studied in this area

What fraction of the distance between where progress was when you started working in the area (A) and where it would need to be to attain human level abilities in the area (C) have we come so far (B)?

%

Divide the period you have worked in the area into two halves: the first and the second.

In which half was the rate of progress in your area higher?

- The first half
- The second half
- They were about the same

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Meta and sociology

4 of 7

To what extent do you think you disagree with the typical AI researcher about when HLMI will exist?

- A lot
- A moderate amount
- Not much

If you disagree, why do you think that is?

To what extent do you think people's concerns about future risks from AI are due to misunderstandings of AI research?

- Almost entirely
- To a large extent
- Somewhat
- Not much
- Hardly at all

What do you think are the most important misunderstandings, if there are any?

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Speed and safety

What rate of global AI progress over the next five years would make you feel most optimistic for humanity's future? Assume any change in speed affects all projects equally.

- | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Much slower | Somewhat slower | Current speed | Somewhat faster | Much faster | Other |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

20 year AI traits

In 2043, how likely do you think the following will be for at least some state-of-the-art AI systems?

	Very unlikely (<10%)	Unlikely (10-40%)	Even chance (40-60%)	Likely (60%-90%)	Very likely (>90%)	I don't understand
Sometimes deceive humans to achieve a goal without this being intended by humans	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Can be trusted to accurately explain their actions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Frequently behave in ways that are surprising to humans	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Make design improvements to increase their own performance regardless of human wishes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Can talk like an expert human on most topics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Take actions						

achieve goals |

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Tasks - fixed probabilities

5 of 7

How many years until you think the following AI tasks will be feasible with:

- a small chance (10%)?
- an even chance (50%)?
- a high chance (90%)?

Let a task be '**feasible**' if one of the best resourced labs could

implement it in less than a year if they chose to. Ignore the question of whether they would choose to.

Tasks

Translate a text written in a newly discovered language into English as well as a team of human experts, using a single other document in both languages (like a Rosetta stone). Suppose all of the words in the text can be found in the translated document, and that the language is a difficult one.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Translate speech in a new language given only unlimited films with subtitles in the new language. Suppose the system has access to training data for *other* languages, of the kind used now (e.g. same text in two languages for many languages and films with subtitles in many languages).

small chance (10%) years

even chance (50%) years

high chance (90%) years

Perform translation about as good as a human who is fluent in both languages but unskilled at translation, for most types of text, and for most popular languages (including languages that are known to be difficult, like Czech, Chinese and Arabic).

small chance (10%) years

even chance (50%) years

high chance (90%) years

Provide phone banking services as well as human operators can, without annoying customers more than humans. This includes many one-off tasks, such as helping to order a replacement bank card or clarifying how to use part of the bank website to a customer.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Correctly group images of previously unseen objects into classes, after training on a similar labeled dataset containing completely different classes. The classes should be similar to the ImageNet classes.

small chance (10%) years

even chance (50%) years

high chance (90%) years

One-shot learning: see only one labeled image of a new object, and then be able to recognize the object in real world scenes, to the extent that a typical human can (i.e. including in a wide variety of settings). For example, see only one image of a platypus, and then be able to recognize platypuses in nature photos. The system may train on labeled images of other objects.

Currently, deep networks often need hundreds of examples in classification tasks¹, but there has been work on one-shot learning for both classification² and generative tasks³.

¹ Lake et al. (2015). Building Machines That Learn and Think Like People

² Koch (2015). Siamese Neural Networks for One-Shot Image Recognition

³ Rezende et al. (2016). One-Shot Generalization in Deep Generative Models

small chance (10%) years

even chance (50%) years

high chance (90%) years

See a short video of a scene, and then be able to construct a 3D model of the scene good enough to create a realistic video of the same scene from a substantially different angle.

For example, constructing a short video of walking through a house from a video taking a very different path through the house.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Transcribe human speech with a variety of accents in a noisy environment as well as a typical human can.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Take a written passage and output a recording that can't be distinguished from a voice actor, by an expert listener.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Routinely and autonomously prove mathematical theorems that are publishable in top mathematics journals today, including generating the theorems to prove.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Perform as well as the best human entrants in the Putnam competition—a math contest whose questions have known solutions, but which are difficult for the best young mathematicians.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Defeat the best Go players, training only on as many games as the best Go players have played.

For reference, DeepMind's AlphaGo has probably played a hundred million games of self-play, while Lee Sedol has probably played 50,000 games in his life¹.

¹ Lake et al. (2015). Building Machines That Learn and Think Like People

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Beat the best human Starcraft 2 players at least 50% of the time, given a video of the screen.

Starcraft 2 is a real time strategy game characterized by:

- Continuous time play
- Huge action space
- Partial observability of enemies
- Long term strategic play, e.g. preparing for and then hiding surprise attacks.

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Play a randomly selected computer game, including difficult ones, about as well as a human novice, after playing the game less than 10 minutes of game time. The system may train on other games.

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Play new levels of Angry Birds better than the best human players. Angry Birds is a game where players try to efficiently destroy 2D block towers with a catapult. For context, this is the goal of the IJCAI Angry Birds AI competition¹.

¹ aibirds.org

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Outperform professional game testers on all Atari games using no game-specific knowledge. This includes games like Frostbite, which require planning to achieve sub-goals and initially posed problems for deep Q-networks^{1, 2}.

¹ Mnih et al. (2015). Human-level control through deep reinforcement learning

² Lake et al. (2015). Building Machines That Learn and Think Like People

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Outperform human novices on 50% of Atari games after only 20 minutes of training play time and no game specific knowledge.

For context, the original Atari playing deep Q-network outperforms professional game testers on 47% of games¹, but used hundreds of hours of play to train².

¹ Mnih et al. (2015). Human-level control through deep reinforcement learning

² Lake et al. (2015). Building Machines That Learn and Think Like People

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Fold laundry as well and as fast as the median human clothing store employee.

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Beat the fastest human runners in a 5 kilometer race through city streets using a bipedal robot body.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Physically assemble any LEGO set given the pieces and instructions, using non-specialized robotics hardware.

For context, Fu 2016¹ successfully joins single large LEGO pieces using model based reinforcement learning and online adaptation.

¹ Fu et al. (2016). One-Shot Learning of Manipulation Skills with Online Dynamics Adaptation and Neural Network Priors

small chance (10%) years

even chance (50%) years

high chance (90%) years

Learn to efficiently sort lists of numbers much larger than in any training set used, the way Neural GPUs can do for addition¹, but without being given the form of the solution.

For context, the original Neural Turing Machines could not do this², but Neural Programmer-Interpreters³ have been able to do this by training on stack traces (which contain a lot of information about the form of the solution).

1 Kaiser & Sutskever (2015). Neural GPUs Learn Algorithms

2 Zaremba & Sutskever (2015). Reinforcement Learning Neural Turing Machines

3 Reed & de Freitas (2015). Neural Programmer-Interpreters

small chance (10%)	<input type="text"/>	years
even chance (50%)	<input type="text"/>	years
high chance (90%)	<input type="text"/>	years

Write concise, efficient, human-readable Python code to implement simple algorithms like quicksort. That is, the system should write code that sorts a list, rather than just being able to sort lists.

Suppose the system is given only:

- A specification of what counts as a sorted list
- Several examples of lists undergoing sorting by quicksort

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Answer any “easily Googleable” **factoid** questions posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet.

Examples of factoid questions:

- “What is the poisonous substance in Oleander plants?”
- “How many species of lizard can be found in Great Britain?”

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Answer any “easily Googleable” factual but open ended question posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet.

Examples of open ended questions:

- “What does it mean if my lights dim when I turn on the microwave?”
- “When does home insurance cover roof replacement?”

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Give good answers in natural language to factual questions posed in natural language for which there are no definite correct answers.

For example: "What causes the demographic transition?", "Is the thylacine extinct?", "How safe is seeing a chiropractor?"

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Write an essay for a high-school history class that would receive high grades and pass plagiarism detectors.

For example answer a question like 'How did the whaling industry affect the industrial revolution?'

small chance (10%) years

even chance (50%) years

high chance (90%) years

Compose a song that is good enough to reach the US Top 40. The system should output the complete song as an audio file.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Produce a song that is indistinguishable from a new song by a particular artist, e.g. a song that experienced listeners can't distinguish from a new song by Taylor Swift.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Write a novel or short story good enough to make it to the New York Times best-seller list.

small chance (10%) years

even chance (50%) years

high chance (90%) years

For any computer game that can be played well by a machine, explain the machine's choice of moves in a way that feels concise and complete to a layman.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Play poker well enough to win the World Series of Poker.

small chance (10%) years

even chance (50%) years

high chance (90%) years

After spending time in a virtual world, output the differential equations governing that world in symbolic form.

For example, the agent is placed in a game engine where Newtonian mechanics holds exactly and the agent is then able to conduct experiments with a ball and output Newton's laws of motion.

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Given a one-sentence description of the task and given the same information you would give a human to perform this task (such as information about the house), physically install the electrical wiring in a new home, without more input from humans.

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Given a one-sentence description of the task, download and fine-tune an existing open source LLM, without more input from humans. The fine-tune must improve the performance of the LLM on some predetermined benchmark metric.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Given a list of long-standing unsolved problems in mathematics, such as the Millennium Prize problems or one the problems in "Unsolved Problems on Mathematics for the 21st Century", solve one without more input from humans.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Given a one sentence description of a research question in machine learning, conduct a study that would inform the answer to that question and write a paper of a quality that could be accepted at a leading machine learning conference, without more input from humans.

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Given a study published at a leading machine learning conference, replicate the study without more input from humans. The replication must meet the standards of the ML Reproducibility Challenge (<https://paperswithcode.com/rc2022>).

small chance (10%)

years

even chance (50%)

years

high chance (90%)

years

Given a set of specifications, build a website from scratch that can handle payment processing, including the frontend, backend, and secure payment integration, without more input from humans.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Given a one-sentence description of the task and no more input from humans, find and patch a security flaw in an open source project with over 100,000 users.

small chance (10%) years

even chance (50%) years

high chance (90%) years

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Tasks - fixed years

5 of 7

How likely do you think it is that the following AI tasks will be feasible within the next:

- 10 years?
- 20 years?
- 50 years?

Let a task be '**feasible**' if one of the best resourced labs could implement it in less than a year if they chose to. Ignore the question of whether they would choose to.

Tasks

Translate a text written in a newly discovered language into English as well as a team of human experts, using a single other document in both languages (like a Rosetta stone). Suppose all of the words in the text can be found in the translated document, and that the language is a difficult one.

10 years % chance

20 years % chance

50 years % chance

Translate speech in a new language given only unlimited films with subtitles in the new language. Suppose the system has access to training data for *other* languages, of the kind used now (e.g. same text in two languages for many languages and films with subtitles in many languages).

10 years % chance

20 years % chance

50 years % chance

Perform translation about as good as a human who is fluent in both languages but unskilled at translation, for most types of text, and for most popular languages (including languages that are known to be difficult, like Czech, Chinese and Arabic).

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Provide phone banking services as well as human operators can, without annoying customers more than humans. This includes many one-off tasks, such as helping to order a replacement bank card or clarifying how to use part of the bank website to a customer.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Correctly group images of previously unseen objects into classes, after training on a similar labeled dataset containing completely different classes. The classes should be similar to the ImageNet classes.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

One-shot learning: see only one labeled image of a new object, and then be able to recognize the object in real world scenes, to the extent that a typical human can (i.e. including in a wide variety of settings). For example, see only one image of a platypus, and then be able to recognize platypuses in nature photos. The system may train on labeled images of other objects.

Currently, deep networks often need hundreds of examples in classification tasks¹, but there has been work on one-shot learning for both classification² and generative tasks³.

¹ Lake et al. (2015). Building Machines That Learn and Think Like People

² Koch (2015). Siamese Neural Networks for One-Shot Image Recognition

³ Rezende et al. (2016). One-Shot Generalization in Deep Generative Models

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

See a short video of a scene, and then be able to construct a 3D model of the scene that is good enough to create a realistic video of the same scene from a substantially different angle.

For example, constructing a short video of walking through a house from a video taking a very different path through the house.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Transcribe human speech with a variety of accents in a noisy environment as well as a typical human can.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Take a written passage and output a recording that can't be distinguished from a voice actor, by an expert listener.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Routinely and autonomously prove mathematical theorems that are publishable in top mathematics journals today, including generating the theorems to prove.

10 years % chance

20 years % chance

50 years % chance

Perform as well as the best human entrants in the Putnam competition—a math contest whose questions have known solutions, but which are difficult for the best young mathematicians.

10 years % chance

20 years % chance

50 years % chance

Defeat the best Go players, training only on as many games as the best Go players have played.

For reference, DeepMind's AlphaGo has probably played a hundred million games of self-play, while Lee Sedol has probably played 50,000 games in his life¹.

¹ Lake et al. (2015). Building Machines That Learn and Think Like People

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Beat the best human Starcraft 2 players at least 50% of the time, given a video of the screen.

Starcraft 2 is a real time strategy game characterized by:

- Continuous time play
- Huge action space
- Partial observability of enemies
- Long term strategic play, e.g. preparing for and then hiding surprise attacks.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Play a randomly selected computer game, including difficult ones, about as well as a human novice, after playing the game less than 10 minutes of game time. The system may train on other games.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Play new levels of Angry Birds better than the best human players. Angry Birds is a game where players try to efficiently destroy 2D block towers with a catapult. For context, this is the goal of the IJCAI Angry Birds AI competition¹.

¹ aibirds.org

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Outperform professional game testers on all Atari games using no game-specific knowledge. This includes games like Frostbite, which require planning to achieve sub-goals and initially posed problems for deep Q-networks^{1, 2}.

¹ Mnih et al. (2015). Human-level control through deep reinforcement learning

² Lake et al. (2015). Building Machines That Learn and Think Like People

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Outperform human novices on 50% of Atari games after only 20 minutes of training play time and no game specific knowledge.

For context, the original Atari playing deep Q-network outperforms professional game testers on 47% of games¹, but used hundreds of hours of play to train².

¹ Mnih et al. (2015). Human-level control through deep reinforcement learning

² Lake et al. (2015). Building Machines That Learn and Think Like People

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Fold laundry as well and as fast as the median human clothing store employee.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Beat the fastest human runners in a 5 kilometer race through city streets using a bipedal robot body.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Physically assemble any LEGO set given the pieces and instructions, using non-specialized robotics hardware.

For context, Fu 2016¹ successfully joins single large LEGO pieces using model based reinforcement learning and online adaptation.

¹ Fu et al. (2016). One-Shot Learning of Manipulation Skills with Online Dynamics Adaptation and Neural Network Priors

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Learn to efficiently sort lists of numbers much larger than in any training set used, the way Neural GPUs can do for addition¹, but without being given the form of the solution.

For context, the original Neural Turing Machines could not do this², but Neural Programmer-Interpreters³ have been able to do this by training on stack traces (which contain a lot of information about the form of the solution).

1 Kaiser & Sutskever (2015). Neural GPUs Learn Algorithms

2 Zaremba & Sutskever (2015). Reinforcement Learning Neural Turing Machines

3 Reed & de Freitas (2015). Neural Programmer-Interpreters

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Write concise, efficient, human-readable Python code to implement simple algorithms like quicksort. That is, the system should write code that sorts a list, rather than just being able to sort lists.

Suppose the system is given only:

- A specification of what counts as a sorted list
- Several examples of lists undergoing sorting by quicksort

10 years

% chance

20 years

% chance

50 years

% chance

Answer any “easily Googleable” **factoid** questions posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet.

Examples of factoid questions:

- “What is the poisonous substance in Oleander plants?”
- “How many species of lizard can be found in Great Britain?”

10 years

% chance

20 years

% chance

50 years

% chance

Answer any “easily Googleable” factual but open ended question posed in natural language better than an expert on the relevant topic (with internet access), having found the answers on the internet.

Examples of open ended questions:

- “What does it mean if my lights dim when I turn on the microwave?”
- “When does home insurance cover roof replacement?”

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Give good answers in natural language to factual questions posed in natural language for which there are no definite correct answers.

For example: "What causes the demographic transition?", "Is the thylacine extinct?", "How safe is seeing a chiropractor?"

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Write an essay for a high-school history class that would receive high grades and pass plagiarism detectors.

For example answer a question like 'How did the whaling industry affect the industrial revolution?'

10 years % chance

20 years % chance

50 years % chance

Compose a song that is good enough to reach the US Top 40. The system should output the complete song as an audio file.

10 years % chance

20 years % chance

50 years % chance

Produce a song that is indistinguishable from a new song by a particular artist, e.g. a song that experienced listeners can't distinguish from a new song by Taylor Swift.

10 years % chance

20 years % chance

50 years % chance

Write a novel or short story good enough to make it to the New York Times best-seller list.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

For any computer game that can be played well by a machine, explain the machine's choice of moves in a way that feels concise and complete to a layman.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Play poker well enough to win the World Series of Poker.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

After spending time in a virtual world, output the differential equations governing that world in symbolic form.

For example, the agent is placed in a game engine where Newtonian mechanics holds exactly and the agent is then able to conduct experiments with a ball and output Newton's laws of motion.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Given a one sentence description of a research question in machine learning, conduct a study that would inform the answer to that question and write a paper of a quality that could be accepted at a leading machine learning conference, without more input from humans.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Given a list of long-standing unsolved problems in mathematics, such as the Millennium Prize problems or one the problems in "Unsolved Problems on Mathematics for the 21st Century", solve one without more input from humans.

10 years % chance

20 years % chance

50 years % chance

Given a one-sentence description of the task and given the same information you would give a human to perform this task (such as information about the house), physically install the electrical wiring in a new home, without more input from humans.

10 years % chance

20 years % chance

50 years % chance

Given a study published at a leading machine learning conference, replicate the study without more input from humans. The replication must meet the standards of the ML Reproducibility Challenge (<https://paperswithcode.com/rc2022>).

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Given a one-sentence description of the task, download and fine-tune an existing open source LLM, without more input from humans. The fine-tune must improve the performance of the LLM on some predetermined benchmark metric.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Given a set of specifications, build a website from scratch that can handle payment processing, including the frontend, backend, and secure payment integration, without more input from humans.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Given a one-sentence description of the task and no more input from humans, find and patch a security flaw in an open source project with over 100,000 users.

10 years	<input type="text"/>	% chance
20 years	<input type="text"/>	% chance
50 years	<input type="text"/>	% chance

Do you have any comments on your interpretation of these questions? (optional)

Which considerations were important in your answers to these questions? (optional)

Safety quote

6 of 7

Stuart Russell summarizes an argument for why highly advanced AI

might pose a risk as follows:

The primary concern [with highly advanced AI] is not spooky emergent consciousness but simply the ability to make high-quality decisions. Here, quality refers to the expected outcome utility of actions taken [...]. Now we have a problem:

- 1. The utility function may not be perfectly aligned with the values of the human race, which are (at best) very difficult to pin down.*
- 2. Any sufficiently capable intelligent system will prefer to ensure its own continued existence and to acquire physical and computational resources – not for their own sake, but to succeed in its assigned task.*

A system that is optimizing a function of n variables, where the objective depends on a subset of size $k < n$, will often set the remaining unconstrained variables to extreme values; if one of those unconstrained variables is actually something we care about, the solution found may be highly undesirable. This is essentially the old story of the genie in the lamp, or the sorcerer's apprentice, or King Midas: you get exactly what you ask for, not what you want.

Do you think this argument points at an important problem?

No, not a real problem.



No, not an important problem.



Yes, a moderately important problem.



Yes, a very important problem.



Yes, among the most important problems in the field.



How valuable is it to work on this problem **today**, compared to other problems in AI?

Much less valuable



Less valuable



As valuable as other problems



More valuable



Much more valuable



How hard do you think this problem is compared to other problems in AI?

Much easier



Easier



As hard as other problems



Harder



Much harder



Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Safety resources + ethics

6 of 7

Let '**AI safety research**' include any AI-related research that, rather than being primarily aimed at improving the *capabilities* of AI systems, is instead primarily aimed at *minimizing potential risks* of AI systems (beyond what is already accomplished for those goals by increasing AI system capabilities).

Examples of AI safety research might include:

- Improving the human-interpretability of machine learning algorithms for the purpose of improving the safety and robustness of AI systems, not focused on improving AI capabilities
- Research on long-term existential risks from AI systems

- AI-specific formal verification research
- Developing methodologies to identify, measure, and mitigate biases in AI models to ensure fair and ethical decision-making.
- Policy research about how to maximize the public benefits of AI

How much should society prioritize **AI safety research**, relative to how much it is currently prioritized?

Much less	Less	About the same	More	Much more
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answer to this question? (optional)

Safety resources

6 of 7

Let '**AI safety research**' include any AI-related research that, rather than being primarily aimed at improving the *capabilities* of AI systems, is instead primarily aimed at *minimizing potential risks* of AI systems (beyond what is already accomplished for those goals by increasing AI system capabilities).

Examples of AI safety research might include:

- Improving the human-interpretability of machine learning algorithms for the purpose of improving the safety and robustness of AI systems, not focused on improving AI capabilities
- Research on long-term existential risks from AI systems
- AI-specific formal verification research
- Policy research about how to maximize the public benefits of AI

How much should society prioritize **AI safety research**, relative to how much it is currently prioritized?

Much less



Less



About the
same



More



Much more



Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answer to this question? (optional)

Extinction

What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species?

. %

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Extinction 2

What probability do you put on future AI advances causing human extinction or similarly permanent and severe disempowerment of the human species within the next 100 years?

. %

What probability do you put on human inability to control future advanced AI systems causing human extinction or similarly permanent and severe disempowerment of the human species?

. %

Do you have any comments on your interpretation of this question? (optional)

Which considerations were important in your answers to this question? (optional)

Demographics

How much thought have you given in the past to **when HLMI (or something similar) will be developed?**

- Very little. e.g. "I can't remember thinking about this."
- A little. e.g. "It has come up in conversation a few times"
- A moderate amount. e.g. "I read something about it now and again"
- A lot. e.g. "I have thought enough to have my own views on the topic"
- A great deal. e.g. "This has been a particular interest of mine"

How much thought have you given in the past to **social impacts of smarter-than-human machines?**

- Very little. e.g. "I can't remember thinking about this."
- A little. e.g. "It has come up in conversation a few times"
- A moderate amount. e.g. "I read something about it now and again"
- A lot. e.g. "I have thought enough to have my own views on the topic"
- A great deal. e.g. "This has been a particular interest of mine"

Are you an AI researcher?

Yes

No

What are your main areas of research?

Where do you work?

Industry

Academia

Other

In which region did you complete your undergraduate study?

- Asia
- North America
- South America
- Africa
- Oceania
- Europe
- Other
- Don't want to say

End 2023 B

Thank you for contributing to the 2023 Expert Survey on Progress in AI!

You should hear from us soon about receiving \$50 or directing it to a charity of your choice.

We hope to send you output from this research as it becomes available.

We are a group of researchers interested in measuring and understanding AI progress and its implications. If you are interested in learning more about this research, please click [here](#) or [email us](#).

Katja Grace,
Lead researcher, AI Impacts

Would you like to be recognized in print as an expert participant in this survey?

Yes No

If you have any questions or comments for us, please feel free to share them below.

If you had any issues with completing this survey, please reach out to harlan@aiimpacts.org.