STAT 165 Final Project

Samuel Bobick* and Zachary Mackin*

May 2024

Abstract

We construct a forecast for the year when the human population on Earth will drop below 400 million. First, we define precise resolution criteria. We conduct a literature review of 3 similar forecasting projects. Then, we enumerate all possible ways the question could resolve with the mutually exclusive and collectively exhaustive (MECE) principle. For each possibility, we assign each event a probability that evolves with time as physical, technological and societal conditions change. To construct these probabilities we rely on background research and a suite of forecasting techniques such as Fermi estimation, combining forecasts, trend extrapolation, and applying common probability distributions. Finally, we use these probabilities as parameters to a Monte Carlo simulation to generate a confidence interval.

1 Background

The sun has about 5 billion to 6 billion years left before it dies and wipes out life on Earth. But will humans perish or leave Earth before that? This could be precipitated by many factors, such as war, famine, natural disaster, or technological advancement leading to space colonies. In our project we will quantify the diverse set of possibilities that lead to Earth's population falling below 400 million people.

2 The Question

When will the human population of planet Earth be less than or equal to 400 million people? If the human population does not drop to 400 million by the year 6 billion AD, then the question will resolve to 6 billion AD.

3 Relevant Resolution Criteria

It is difficult to point to specific resolution criteria for this question. If there is a mass extinction event it is likely that most sources of information will no longer be up to date. As such, we assume that the human population is a directly observable characteristic. This follows the precedent of other mass extinction

*College of Computing, Data Science, and Society

forecasting questions such as the Ragnarok Series [Metaculus, 2024]. Humans are defined by the following criteria:

1. Humans are defined strictly by the scientific definition of *Homo sapiens* [Stringer, 2016]. Humans who are interbred with other species, including aliens, do not count.

2. Humans who take up permanent residence on other planets do not count, even if they visit Earth.

3. If Earth ceases to exist at some point in time due to an astronomical event, then the question resolves to that point in time.

4. Humans who have been augmented with mechanical body parts (e.g. prosthetic limbs, pacemaker, artificial heart) will count in the total. However, humans who have a mechanical or electrical brain augmentation that they cannot live without do not count as humans. This addresses the question of cyborgs.

4 Literature Review

The Ragnarok forecasting series is a series of forecasting questions regarding potential existential risks for humanity [Metaculus, 2024]. These forecasts are segmented by different outcomes, from climate change and environmental collapse to geopolitical conflicts or technological disasters. As such, they provide useful inspiration for our construction of possibilities through the MECE principle. That being said, they do not consider extinction events as dire as our resolution criteria nor at the wide timescale that we consider; most focus on risks within the next century.

Two other texts we utilized were *12 Risks with Infinite Impact* and *Forecasting Existential Risks: Evidence from a Long-Run Forecasting Tournament*, which not only enumerate all of the possibilities for extinction events, but also quantify the risk for extinction events [Karger et al., 2023][Pamlin and Armstrong, 2015]. The first source quantified twelve global challenges that they believe encompassed all possibilities for something with "infinite risk" (i.e. an extinction event). This source was helpful as it conglomerated various available assessments and break them down into possible events that could lead to an extinction event. The second source compared extinction predictions of superforecasters to domain experts, and assessed that domain experts tend to overemphasize the risk of their domain, something that was a helpful analysis for us to consider throughout our research [Pamlin and Armstrong, 2015]. The tournament also had forecasters quantify the base rates that could lead to an event, such as quantifying the probability of superhuman artificial intelligence within the AI category. This was helpful to consider both all the possibilities that could create sufficient and/or necessary conditions for an extinction event, but also to inform forecasts of these contributing factors.

5 Considering Possibilities with the MECE Principle

In this section we apply the MECE principle to determine all possible factors that could result in human extinction or relocation from Earth.

5.1 Natural Disaster

In all of world history that there is geological evidence for, we identify one geological event that reduced human population drastically enough to satisfy our criteria – the Toba Catastrophe [Dhar, 2022]. The Toba catastrophe killed an estimated 97% of humans, caused global cooling which affected plant growth, and may have triggered another supervolcano eruption on the other side of the world. We use the Toba catastrophe as a reference class in our base rate esimation in Section 8.1.

In our research, we considered other natural disasters such as storms, earthquakes, and tsunamis, but determined that the localized nature of these events make it essentially impossible for such an event to cause the mass extinction we require.

5.2 Astronomical Events

A massive meteorite, asteroid, or comet strike such as the Chicxulub asteroid 65 million years ago could cause human extinction though direct impact and ensuing global cooling. We can use the asteroid crater record as a reference class [Ear, 2024]. We lump these events together as "space rock strike" in our simulation setup in Section 6.

Other astronomical events that could cause a mass human extinction include a gamma ray burst, supernova, or random encounter, where a black hole, star, or other astronomical body enters the solar system [Betz, 2023]. Due to the astronomical magnitude of these events, humans have little hope of surviving one of them, even with advanced technology. In our simulation we lump these events together as "random encounter" in our simulation setup in Section 6.

In about 1.3 billion years, the sun will enter its next phase of life and begin to burn much hotter [Stringer, 2016]. At this point, the Earth will certainly be uninhabitable. Models suggest that the oceans will evaporate and biologial life forms will die out.

5.3 Climate Change

In our research, we consider both manmade and natural climate change. However, we find the drastic population reduction required by our resolution criteria as highly unlikely to be affected by climate change. While climate change can affect human resources such as crop yields, livable coastal land, and desirable temperatures, we concur with experts that climate change will almost surely not be a cause of mass human extinction [MIT Climate, 2023].

5.4 Biological Considerations

We consider two different scenarios for a disease to pose an existential threat to humanity. This includes a naturally occurring pathogen and a human-constructed or modified pathogen. Note that the manmade pathogen could be either intentionally or unintentionally released.

Both of these may pose an existential threat to humanity. However, most diseases that are deadly enough to kill that many humans run the risk of developing less deadly mutations. Even the worst pandemic in the modern era only killed between $1 - 5.4\%$ of the world's population [Roser, 2020]. However, we cannot count out this scenario due to potential improvements in genetic modification technology.

5.5 Societal Considerations

Since July 16th, 1945, when the Trinity test demonstrated proof of concept of the atomic bomb, humanity has had the capability of hitting this criteria through nuclear warfare. Throughout the Cold War, many people lived in fear of an all-out nuclear war, with presidential campaign ads even claiming who you voted for could cause nuclear war [Bernbach and Schwartz, 1964]. This risk certainly still exists given the number of nuclear warheads that still exist and the unpredictability of global conflict. We consider the possibility of both intentionally-inflicted nuclear warfare and accidental releases of nuclear warfare.

5.6 Technological Factors

Many technological factors exist to resolve this question. For example, humans may leave Earth to colonize other planets. This could be to escape disaster on Earth or to increase standard of living for humans if Earth's natural resources are exhausted.

We also consider artificial intelligence. Superforecasters predict an AI catastrophe as the most likely existential threat to occur in the next 100 years [Karger et al., 2023]. Though experts vary in their predictions, some fear that creating superhuman AI, something that is certainly possible with the incredibly quick progress of AI, is a Pandora's box that could result in the extinction of humanity. The ways that AI could cause human extinction range from a rogue AI committing some sort of attack that kills most humans to superhuman AI agents deciding that humans are unnecessary or even detrimental to their objectives and wiping them out.

Finally, improvements in technology could lead to biological humans being eradicated in favor of cyborgs, as defined in our resolution criteria. Since this involves augmenting human capabilities with computerized and mechanical intelligence, we lump this outcome into AI catastrophe in our simulation setup in Section 6.

5.7 Other Options

We consider several other options that could achieve our resolution criteria. In general, *homo sapiens* has been the species most capable of controlling its environment. However, there are several ways that this could cease to be the case. One is an alien invasion, where a species that originates from another planet invades Earth and takes humans' control of their environment and natural resources.

Another option is evolution. Just as neanderthals died out to humans, it is possible a new species could emerge and overtake *homo sapiens*. Additionally, there exists minimal genetic diversity within *homo sapiens* with certain tribes of chimpanzees having more genetic diversity than all of humankind [Kaessmann et al., 1999]. This makes humankind more susceptible to evolutionary pressures.

Finally, a difficult to predict threat is one that has yet to emerge. For example, 100 years ago nobody was worried about nuclear warfare or AI. We hypothesize that new, unknown man-made threats to humanity will emerge in the future.

6 Assigning Probabilities to these Scenarios

For each risk R identified in Section 5, we assign a probability P(R occurs in 100 years).

We use a Monte Carlo simulation with 100-year timesteps. At each timestep, each probability evolves according to some update rule to account for changing world conditions.

In general, we employ decaying exponential update rules to model risks that we think have an increasingly smaller probability of causing a mass extinction event as time passe. For example, consider the volcano risk. As technology improves as time passes, we expect humans to be able to better preserve themselves in the case of a massive volcanic eruption. As such, we decrease the probability by 0.001% at each timestep in our Monte Carlo simulation.

On the other hand, exponentially increasing update rules indicate risks that are emerging and will become more serious in the future. For instance, the risk of humans leaving earth to colonize other planets increases exponentially from a risk of 10^{-8} at a rate of 0.001% for every 100 years, since we imagine that as technology progresses, so does human capability to leave Earth for other planets.

The input base probabilities and update rules to our simulation are detailed in Section 1. To generate these probabilities, we apply forecasting techniques detailed in Section 8.

Table 1: Input probabilities to Monte Carlo Simulation

Risk	Probability	Source
Volcano	$5 \times 10^{-7} \times 0.99999^{(x-2000)/100}$	8.1
Space rock strike	$8 \times 10^{-8} \times 0.999^{(x-2000)/100}$	8.2
Random encounter	10^{-9}	[Betz, 2023]
Sun burns hotter	$\mathbb{I}(\text{year} = 1.3 \times 10^{9})$	[Wolf and Toon, 2015]
Natural climate change	$10^{-9} \times 0.9999^{(x-2000)/100}$	[MIT Climate, 2023]
Manmade climate change	$10^{-9} \times 0.9999^{(x-2000)/100}$	[MIT Climate, 2023]
Naturally occurring pathogen	$8 \times 10^{-8} \times 0.9999^{(x-2000)/100}$	8.3
Biological warfare	$\frac{10^{-7}}{1+e^{-0.01(x-2000)}}$	8.4
AI Catastrophe	$\begin{cases} 7 \times 10^{-3} & \text{for } x \leq 2400 \\ \frac{10^{-8}-0.007}{1600} \times x + (0.007 - 2400 \times \frac{10^{-8}-0.007}{1600}) & \text{for } 2400 < x \leq 4000 \\ 10^{-8} & x > 4000 \end{cases}$	8.5 [Karger et al., 2023]
Humans colonize other planets	$10^{-8} \times 1.00001^{(x-2000)/100}$	[Caballero, 2022]
Nuclear Warfare	$5 \times 10^{-6} \times 0.9999^{(x-2000)/100}$	8.6[Pamlin and Armstrong, 2015]
Alien invasion	10^{-11}	[Caballero, 2022]
Evolution	3.6×10^{-7}	8.7
New threat	$8 \times 10^{-7} \times 1.00001^{(x-2000)/100}$	8.8

7 Results and Discussion

Code for our simulation is linked here. To generate our confidence interval we use a Monte Carlo simulation with 1000 iterations. Due to the large temporal scale of our question, the simulation is computationally-intensive and takes several hours to run.

Our final confidence interval is [232,600 AD, 34,814,000 AD]. Figure 1 shows the simulated resolution year distribution. Figure 2 shows the causes of the resolutions.

Qualitatively, we provide the following observations. Firstly, there is a large cluster of resolutions in the first 500,000 years. These mainly correspond to the threats of AI and nuclear warfare. If these risks do not perpetuate soon, our model makes them less likely, so for the simulations that pass those 500,000 years, emerging threats such as a new threat, evolution, or human colonization of other planets begin to begin to take over. Interestingly, our model predicts our question will resolve before changes in the sun's life cycle make Earth uninhabitable.

A new threat caused extinction about 75% of the time. Given the large temporal extent of our forecasting question, we argue that this is reasonable. Several of the most threats of today, such as biological warfare, AI, and nuclear weapons, were unknown a mere 100 years ago. As time progresses, we concur that threats that were unknown to humankind in 2024 will be the most significant factor in resolving our forecasting question.

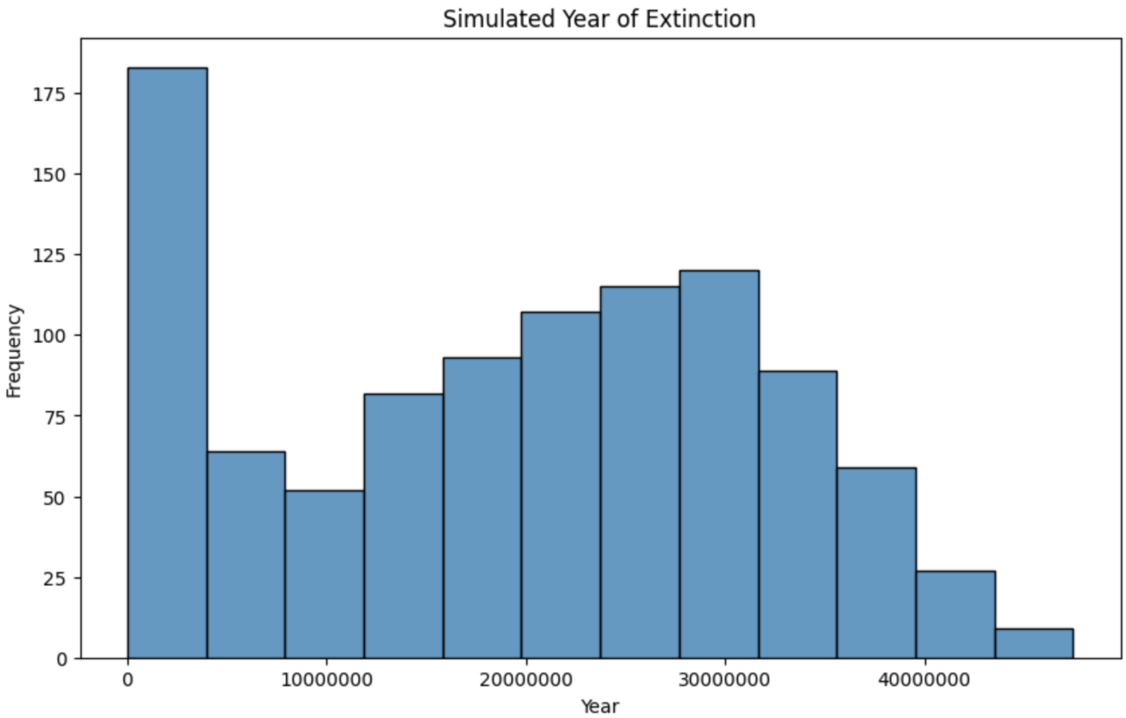

Figure 1: Distribution of year of resolution in our simulation.

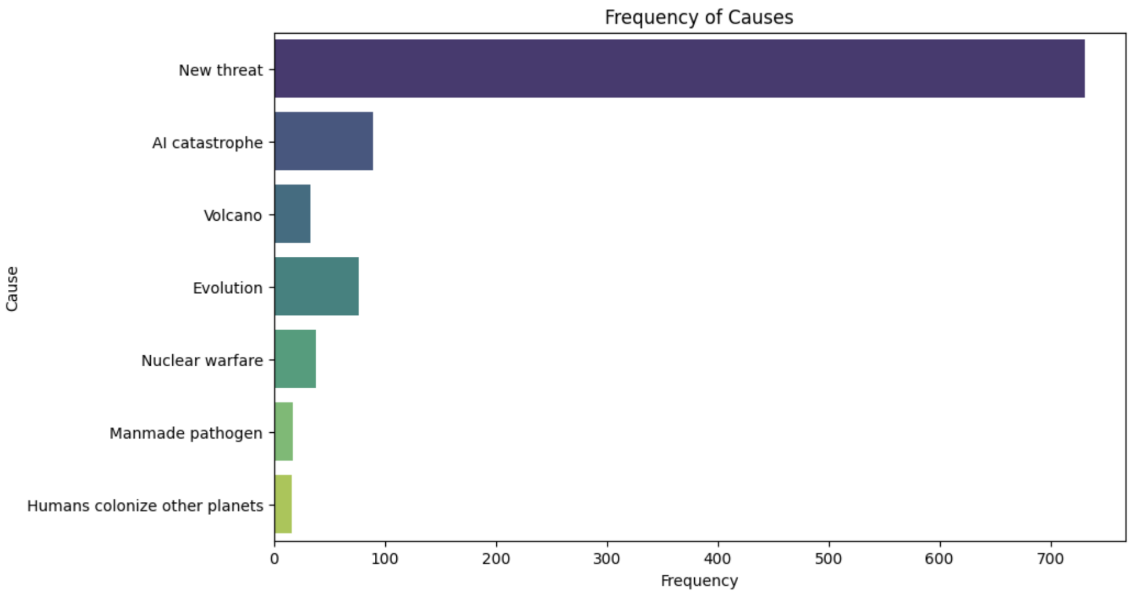

Figure 2: Cause of extinction in our simulation.

8 Appendix: Deriving Simulation Parameters

8.1 Volcano Base Rate

The Toba catastrophe killed 97% of all humans [Rampino and Ambrose, 2000]. The *Homo* genus has existed for 200 million years. As such, we assume a simple base rate of 1 qualifying event per 200 million years. Since we are taking 100-year timesteps, our probability comes out to

$$\frac{1 \text{ deadly event}}{200 \times 10^6 \text{ years}} \times 100 \text{ years per timestep} = 5 \times 10^{-7}$$

8.2 Space Rock Strike Base Rate

We use the Earth impact database for a list of crater impacts on Earth [Ear, 2024]. Meteorite/asteroid/comet strike distributions are known to follow lognormal distributions [Bright et al., 2011]. We fit the crater sizes to the Earth impact database lognormal distribution with MLE. Then, we assume the minimum crater size corresponding to an extinction event is a diameter of 150 km, the size of the Chicxulub crater from the meteorite that killed the dinosaurs. These assumptions give us a base rate of 3.5 qualifying impacts in the the 5 billion year history of Earth. We adjust upward slightly to 4 to account for the fact that there may be a crater that is undiscovered. This yields a base rate of

$$\frac{4 \text{ deadly events}}{5 \times 10^9 \text{ years}} \times 100 \text{ years per timestep} = 8 \times 10^{-8}.$$

8.3 Naturally Occurring Pathogen Base Rate

We use Wikipedia's list of all pandemics known to kill at least 1 million people [Wikipedia, 2024]. Also, we use the fact that hominids have existed for 7 million years. We estimate that, excluding the modern pandemics of COVID-19 and HIV/AIDS, all the death toll of all pandemics would be 80% lower than actualized due to the advent of modern medicine. After transforming the death tolls by the 80% decrease, we fit a lognormal.

$$\frac{4 \text{ deadly events}}{2 \times 10^8 \text{ years}} \times 100 \text{ years per timestep} = 2 \times 10^{-8}.$$

8.4 Biological Warfare Model

We believe the greatest risk of biological warfare will occur at the time when biotech tools such as protein design are cheap and effective due to artificial intelligence. For the long run, we use a median estimate of 10^{-7} which gives the a 100-year probability of bioweapons occurring adjusted for the probability they would lead to an extinction event[Karger et al., 2023]. To model the growth of technology growth we use the following sigmoid function:

$$P(x) = \frac{10^{-7}}{1 + e^{-0.01(x-2000)}}$$

Where x is the year of interest. This model gives us the "growth" segment of the sigmoid function until around 2200, which is broadly consistent with forecasts about when explosive AI breakthroughs will occur.

8.5 Artificial Intelligence Catastrophe Model

Artificial Intelligence is difficult to develop a base rate for, as it is a new technology. Thus, we aggregate various superforecasters' and domain experts' risk in the next 100 years placing 80% weight on the superforecasters and 20% on domain experts, due primarily to the fact that domain experts tend to overestimate the risk of their domain [Karger et al., 2023].

$$0.8 \times (0.38) + 0.2 \times (2.13) = 0.73$$

Ultimately, we decided to adjust downward given the fact that humans appear to overstate the risk of things they do not quite understand [Rosenberg et al., 2024]. Examples of this phenomenon include 20th-century fears of overpopulation, alien invasion, particle accelerators, and many other technological advancements. We estimate this overstatement, for low probability events, is roughly a factor of 10, even amongst experts. This gives us a rough short-term estimate of 7×10^{-2} for the probability of AI-caused extinction over a 100-year time frame. However, we believe this probability will quickly fall off, as the forecasted median date of a super human-level AI generally falls within the period 2046-2060 [Karger et al., 2023]. If this does not cause a human extinction event within the first few 100 year time frames, it is unlikely this technology ever will, thus the piece wise decrease to 10^{-8}. Our final model comes out to

$$P(x) = \begin{cases} 7 \times 10^{-3} & \text{for } x \leq 2400 \\ \frac{10^{-8} - 0.007}{1600} \times x + (0.007 - 2400 \times \frac{10^{-8} - 0.007}{1600}) & \text{for } 2400 < x \leq 4000 \\ 10^{-8} & x > 4000 \end{cases}$$

8.6 Nuclear Weapons Base Rate

For nuclear warfare, we combine forecasts of domain experts and superforecasters who seem to arrive at the conclusion that the risk in the next 100 years is around 5×10^{-3} [Pamlin and Armstrong, 2015]. However, we think these forecasts suffer from a similar tendency to overstate the risk of revolutionary technology, with some forecasters even referencing President John F. Kennedy's assessment that the Cuban Missile Crisis had between a 33-50% chance of causing a nuclear conflict [Allison, 2012]. We believe this, along with estimations of the probability of other "close calls" are significantly overstated, as is the conclusion that the majority of nuclear conflicts would lead to extinction events.

Due to this we adjust the probability to $5 * 10^{-6}$, with a decreased probability by 0.01% at each timestep, not due to modified defense technology, as we imagine nuclear defense technology will always scale at a similar rate as nuclear offense strategy, but rather due to improved technology in preventing nuclear fallout.

Additionally, human society has shown a tendency to become more peaceful over time.

8.7 Evolution Base Rate

Though many species in the Homo genus have gone extinct, we postulate the only ones similar enough to Homo Sapiens to serve as a reference class are *Homo Neanderthalensis*. Given that the *Homo* genus has existed for 280 million years, and considering the time steps we are left with a base rate of

$$\frac{1 \text{ deadly event}}{2 \times 10^8 \text{ years}} \times 100 \text{ years per timestep} = 3.6 \times 10^{-5}$$

Given the resiliency, large population, and intelligence of *Homo Sapiens*, we postulate that humans are an order of 100 times more resilient than *Homo Neanderthalensis*. Thus we are left with 3.6×10^{-7}.

8.8 New Threat Base Rate

Given that a fear of nuclear Bombs or artificial intelligence causing extinction did not exist before, it seems reasonable to conclude that there will be something invented in either future that poses a similar existential threat. This is hard to develop a base rate for, however it seems a reasonable reference class to include these newly developed threats, and put the threat of a new technology causing extinction similar to these, with a slight upward trajectory with the development of new technologies in the future. We arrive at the probability function

$$P(x) = 8 \times 10^{-7} \times 1.00001^{(x-2000)/100}$$

References

[Ear, 2024] (2024). Earth impact database. Available online.

[Allison, 2012] Allison, G. (2012). The cuban missile crisis at 50. *Foreign Affairs*, 91(4). Retrieved 9 July 2012.

[Bernbach and Schwartz, 1964] Bernbach, D. D. and Schwartz, T. (1964). Daisy (advertisement). Television advertisement. Aired as part of Lyndon B. Johnson's 1964 presidential campaign.

[Betz, 2023] Betz, E. (2023). How will life on earth end? *Astronomy.com*.

[Bright et al., 2011] Bright, J., Aylott, M., Bate, S., Geys, H., Jarvis, P., Saul, J., and Vonk, R. (2011). Recommendations on the statistical analysis of the comet assay. *Pharmaceutical statistics*, 10(6):485–493.

[Caballero, 2022] Caballero, A. (2022). Estimating the prevalence of malicious extraterrestrial civilizations.

[Dhar, 2022] Dhar, M. (2022). What was earth's biggest explosion? Accessed on 2024-04-14.

[Kaessmann et al., 1999] Kaessmann, H., Wiebe, V., and Pääbo, S. (1999). Extensive nuclear dna sequence diversity among chimpanzees. *Science*, 286(5442):1159–1162.

[Karger et al., 2023] Karger, E., Rosenberg, J., Jacobs, Z., Hickman, M., Hadshar, R., Gamin, K., Smith, T., Williams, B., McCaslin, T., Thomas, S., and Tetlock, P. E. (2023). Forecasting existential risks: Evidence from a long-run forecasting tournament. Technical report, Forecasting Research Institute.

[Metaculus, 2024] Metaculus (2024).

[MIT Climate, 2023] MIT Climate (Accessed 2023). Will climate change drive humans extinct or destroy civilization? Online.

[Pamlin and Armstrong, 2015] Pamlin, D. and Armstrong, S. (2015). *12 Risks that threaten human civilisation: The case for a new risk category*.

[Rampino and Ambrose, 2000] Rampino, M. R. and Ambrose, S. H. (2000). Volcanic winter in the garden of eden: the toba supereruption and the late pleistocene human population crash. *SPECIAL PAPERS-GEOLOGICAL SOCIETY OF AMERICA*, pages 71–82.

[Rosenberg et al., 2024] Rosenberg, J., Karger, E., Morris, A., Hickman, M., Hadshar, R., Gamin, K., Smith, T., Jacobs, Z., and Tetlock, P. (2024). Roots of disagreement on ai risk: Exploring the potential and pitfalls of adversarial collaboration.

[Roser, 2020] Roser, M. (2020). The spanish flu: The global impact of the largest influenza pandemic in history. *Our World in Data*. https://ourworldindata.org/spanish-flu-largest-influenza-pandemic-in-history.

[Stringer, 2016] Stringer, C. (2016). The origin and evolution of homo sapiens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1698):20150237.

[Wikipedia, 2024] Wikipedia (2024). List of epidemics and pandemics — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=List%20of%20epidemics%20and%20pandemics&oldid=1221506390`. [Online; accessed 01-May-2024].

[Wolf and Toon, 2015] Wolf, E. and Toon, O. (2015). The evolution of habitable climates under the brightening sun. *Journal of Geophysical Research: Atmospheres*, 120(12):5775–5794.