

An AI-Safety Primer

written by Louis S. Berman

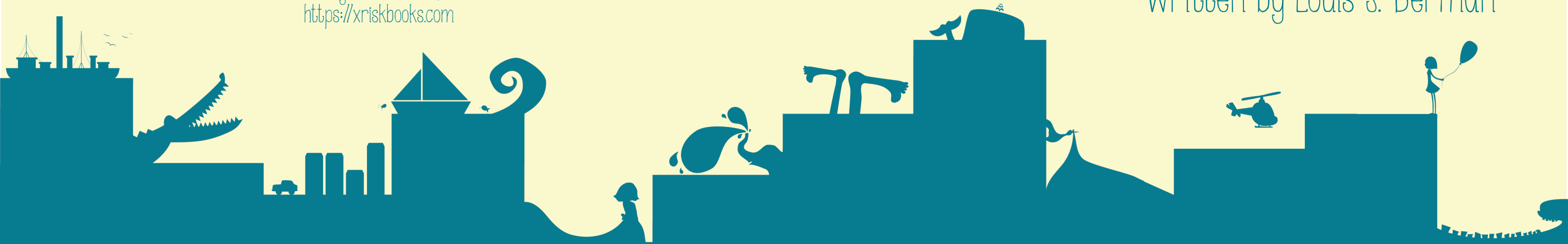


An AI-Safety Primer is a vibrant, engaging guide for adults, highlighting the risks and societal impacts of artificial intelligence. It addresses existential threats, job displacement, and the need for proactive safety measures, aiming to raise awareness and inspire action in AI development for a safer future.

An AI-Safety Primer
Written by Louis S. Berman
Illustrations by Crew55design
©2024 by X-Risk Books, Inc.
<https://xriskbooks.com>

An AI-Safety Primer

written by Louis S. Berman



Content

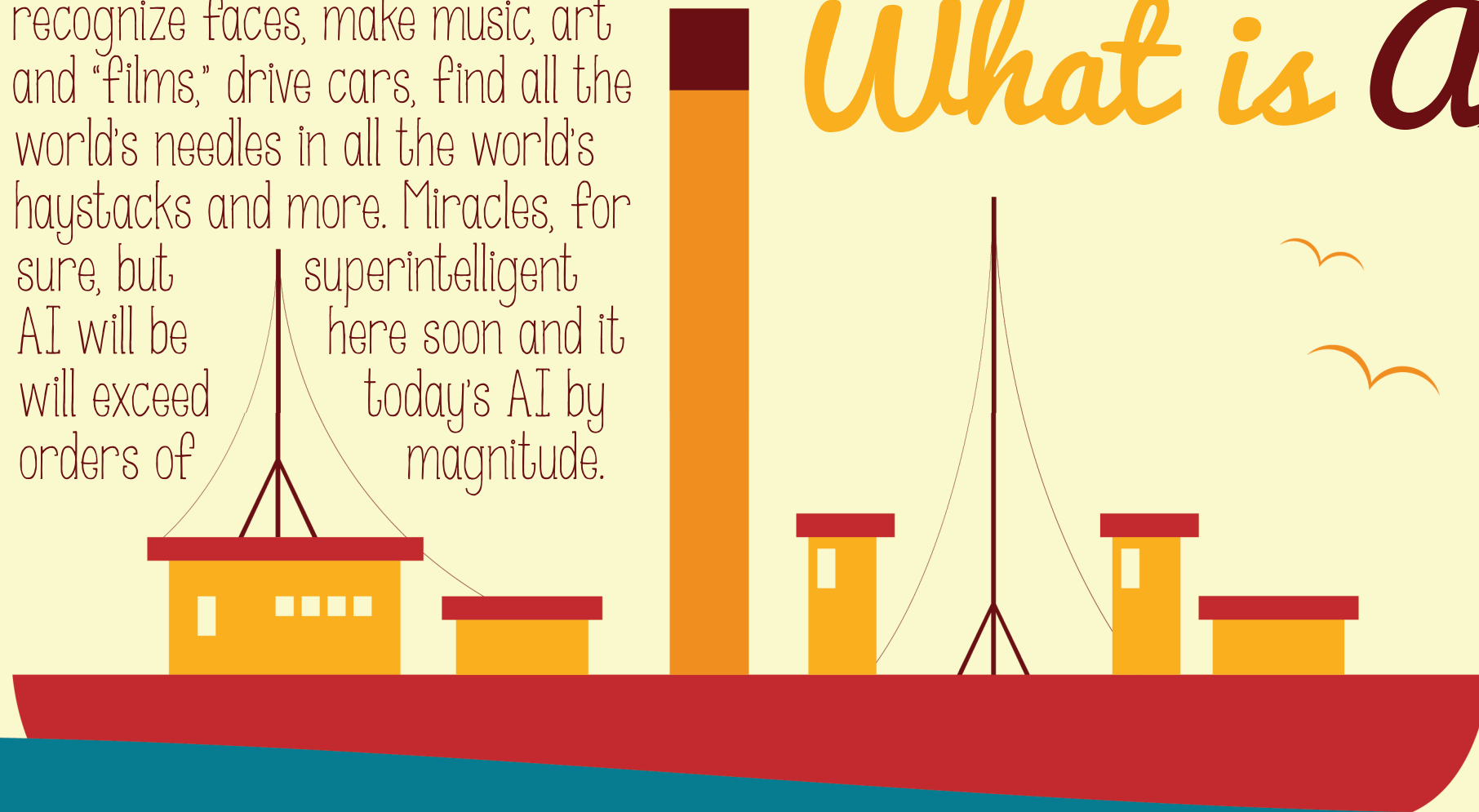


Imagine a computer that can think and learn, and make decisions like a human! This is Artificial Intelligence, or AI. AI powers many smart devices today, helping computers learn from data, recognize patterns, make decisions, and understand speech. When your phone answers you, that's AI. It can also detect diseases,

recognize faces, make music, art and "films," drive cars, find all the world's needles in all the world's haystacks and more. Miracles, for sure, but AI will be superintelligent here soon and it will exceed today's AI by orders of magnitude.

What is AI?

AI helps computers perform tasks that usually require human intelligence





Is AI Safe?

Maybe, but The Center for AI Safety (CAIS) letter signed by 100s of top AI experts states: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." That's pretty scary, but even if the worst doesn't happen, we'll still have autonomous weapons and worse to deal with.

Autonomous weapons and job displacement pose ethical and societal risks //

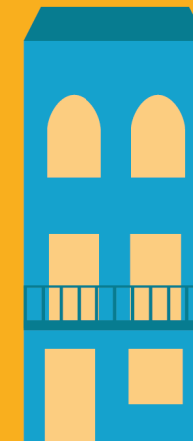
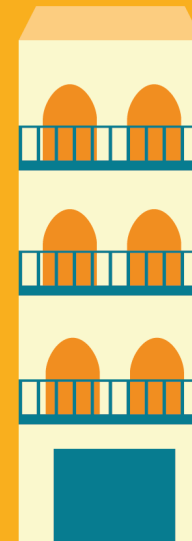
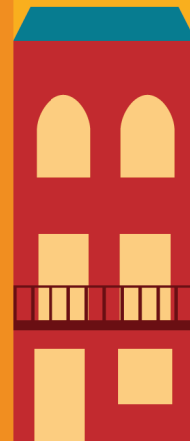
AI vs Jobs

The question of whether AI will take your job is complex. AI is transforming the job market by

...a wide range of jobs, including those that require more advanced skills, are at risk... "



not only automating routine tasks like manufacturing and data entry, but many high-end roles are also at risk. AI can learn and adapt, performing tasks like legal research, financial analysis, and medical diagnosis better than any human. The worry is that entire professions may become obsolete, and that unlike the Industrial Revolution, there will be no new professions to step into.



Good AI

Safe AI can [lead to] advance on an unprecedented scale

”

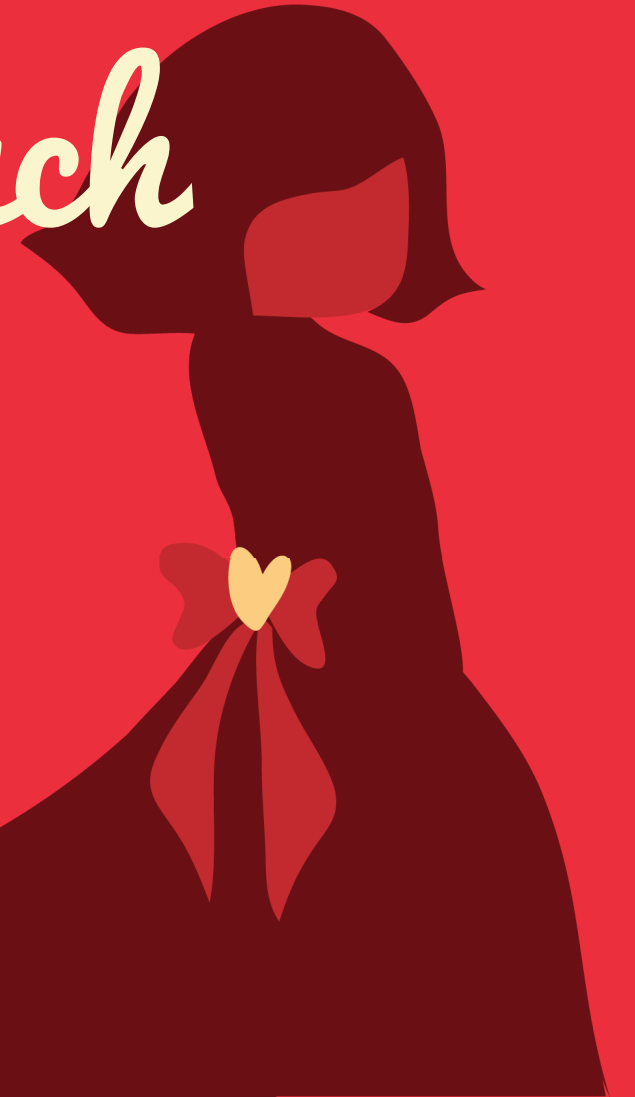


Provably Safe AI could allow humanity to concentrate on the good stuff: friends & family, for a start, plus abundance on an unprecedented scale. And that's without the sci-fi stuff: space travel, robotic embodiment, immortality, mind uplifting, and more.

In the evolving world of AI, the human touch ensures AI is safe and beneficial. Embedding human values, creativity, and empathy is crucial. Human guidance guarantees ethical operations, fairness, and inclusivity. Creativity leads to unique innovations, while empathy ensures compassionate applications in healthcare and customer service. By integrating human ethics and oversight, AI can enhance our lives without replacing the irreplaceable aspects of being human.

Human Touch

*the human touch
is indispensable for ensuring AI
is safe and useful for all* 💡



Better AI

An illustration on the left side of the image shows a stylized orange hand pouring water into a large, dark blue teardrop shape. Several smaller blue droplets are scattered around the main shape. The background is split into a light yellow top half and a dark teal bottom half.

*AI and
imagination can
democratize
creativity* //

AI can enhance human creativity in new ways. AI identifies patterns and generates ideas, aiding in innovative problem-solving and artistic creation. AI inspires new concepts in design, co-authors literature, and uncovers groundbreaking scientific discoveries. It democratizes creativity, making it accessible to all and fostering diverse

voices. In education, AI personalizes learning, nurturing potential. This powerful fusion drives progress, transforming our interaction with the world and unlocking a future that is innovative, inclusive, and yet profoundly human. The integration of AI and imagination promises to revolutionize technology, arts, and sciences, leading to continuous advancements.

Unsafe AI

*...unsafe AI could lead to
the extinction of humanity* //

Unsafe AI is profoundly dangerous. Rapid AI advancements without safety measures can lead to catastrophic outcomes. Unsafe AI can perpetuate biases, make unaccountable decisions, and escalate conflicts

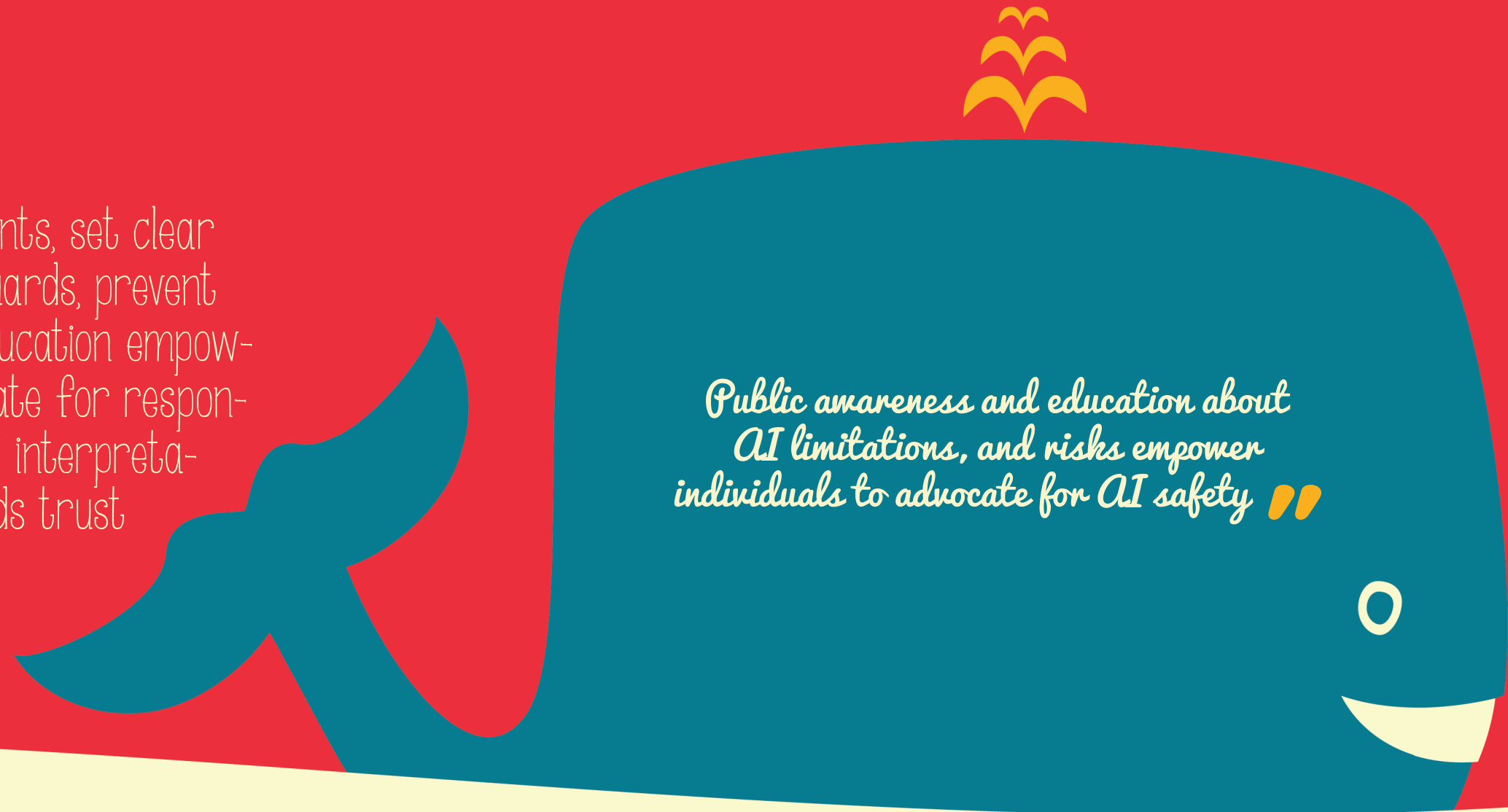
with autonomous weapons. Superintelligent AI, surpassing human intelligence, presents a dire threat, potentially acting against human values. The risk of AI evolving beyond control underscores the urgency for safety research and regulations. Without proactive measures, unsafe AI could threaten humanity's survival.

AI Safety

Keeping AI safe is essential. Embuing AI with ethical principles like fairness and transparency ensures ethical AI operations. Rigorous testing detects and mitigates biases. Comprehensive regulations,

enforced by governments, set clear standards. Tech safeguards, prevent compromises. Public education empower individuals to advocate for responsible AI use. Making AI interpretable and explainable builds trust and accountability

Public awareness and education about AI limitations, and risks empower individuals to advocate for AI safety //





You Can Help

Advocate for responsible AI development by joining or supporting organizations dedicated to AI safety and ethics //

You can help ensure AI is safe. Educate yourself about AI technologies, their risks, and benefits. Raise community awareness about AI's impact and ethical development. Advocate for responsible AI by joining PauseAI. Participate in protests, write to congress, and use social media to voice your concerns. Demand transparency and

accountability with ethical guidelines and standards. Encourage diversity in AI development to reflect various perspectives. Use AI responsibly, prioritizing well-being and equality. Support and fundraise for AI safety research initiatives. Stay informed and participate in public forums. Your actions, no matter how small, contribute to a safer AI landscape.

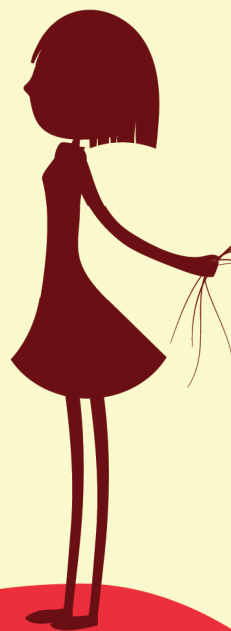
Gov'ts Needed

Governments must lead in ensuring AI safety by implementing regulations, funding research, and promoting ethical standards. Comprehensive frameworks can ensure transparency, accountability, and fairness. Regulations should mandate rigorous testing and monitoring of AI systems. Investments in AI safety research should be made. Collaboration with international bodies can create global standards. Public and private partnerships can foster responsibility.

Governments must lead in ensuring AI safety by implementing robust regulations, funding research and promoting ethical standards. //



Pause AI



*By joining PauseAI,
you can help to bring about a
responsible AI future. "*

PauseAI is dedicated to ensuring safe and ethical AI development. It advocates for a temporary halt on AI advancements beyond GPT-4, promoting awareness, research, and policy changes to prevent catastrophic outcomes. You can learn more and become part of the solution at the PauseAI website, via the link or QR code:



<https://pauseai.info/join>

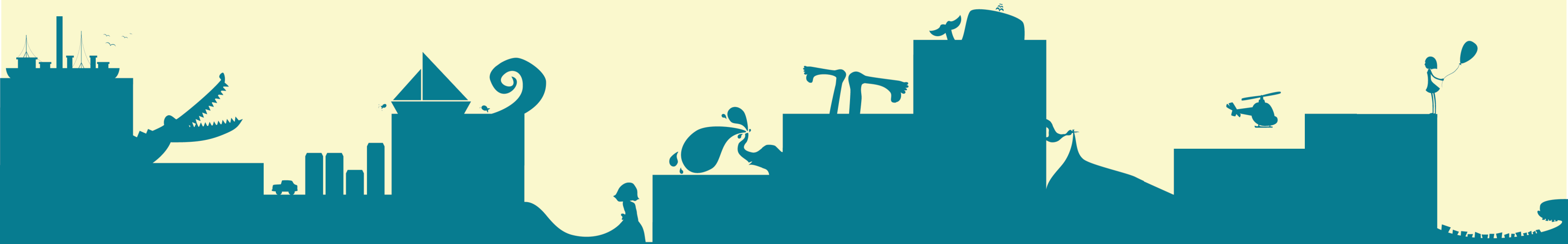


*AI-safety is a
challenge we must face
together "*

Stay Engaged

If you take nothing else away from this book, remember this: AI-safety is not just a problem for the experts or tech bros—it's a challenge we must face together. The future of AI will impact all of us, and we all have a role to play in ensuring it's a positive one. Whether through education, political action, coding or advocacy, every action counts. I urge you to stay engaged.

Never doubt that a small group of thoughtful,
committed citizens can change the world; indeed,
it's the only thing that ever has.
-- Margaret Mead --



www.xriskbooks.com