Boom or Doom

Points to Ponder about Artificial Intelligence

By Ken Ing

We often hear "*A.I. is just computer code. It just does what programmers tell it to*". But now A.I. programmers have told A.I. how to be self-taught. Doesn't this "Deep Learning" mean programmers no longer can predict, let alone control, what A.I. will do?

We often hear "A.I. will never take the place of humans, because it does not experience emotions or have feelings". But can't all emotions and feelings be imitated, making it virtually impossible for us to tell the difference between A.I. and a real person?

Do you agree with this list of goals: Well-adjusted people, nourishing relationships, secure security, engaged communities, and an effective nation? Shouldn't A.I. be judged by its impact on these goals? Are there other goals that could offset the negative impact of A.I. on those 5 goals?

Haidt documents a doubling of instances of self-harm, anxiety, and suicides for young people in the 1st 5 years of smartphones, plus major increases in pessimism, depression, and loneliness, with time spent with friends cut in half. Is the new A.I. technology going to make this worse?

The U.S. Surgeon General says a lack of social connection is more dangerous than smoking 15 cigarettes a day. The U.K. and Japan have appointed "Ministers of Loneliness". Shouldn't the impact on loneliness and deaths of despair be major considerations in our choices about A.I.?

We marvel at how tools like ChatGPT can combine many data sources to produce insightful analysis. But we have no way of knowing what it bases its conclusions on. Should we just trust what it says, or should we be skeptical of everything it says?

Our attachment to grievances, self-aggrandizement, growing mental illnesses, increasing loneliness, and a normalization of political violence are societal ills threatening our body politic. Shouldn't A.I. be judged by whether it improves, or worsens, these cultural trends?

Jonathan Haidt points to smartphones and social media as a major contributor to an alarming deterioration of the mental health of young girls since 2009. Should this be a cautionary tale about our ability to assess the wisdom of rapidly adopting powerful new technologies?

Smartphones and social media are not the only technologies that have loosened our societal binds. Television, Cable TV, and the Internet share the blame for our toxic politics and decay of community. Were we too optimistic about the Internet because we ignored human nature?

Our politics are a mess. Both sides judge that the "other" side is closed minded, dishonest, immoral, and unintelligent. Shouldn't the impact on our political polarization be a major factor in our choices about how, and whether, to adopt new A.I. technologies?

Boom or Doom

Points to Ponder about Artificial Intelligence

By Ken Ing

To achieve our goals, people need to feel they have a purpose, people need to do things together, we need to carefully manage risks, we need a broad sense of who is "us", and we need trust at all levels of society. Are these the proper objectives to test proposed policy decisions about A.I.?

People are falling in love with A.I. Chatbots. A.I. boyfriends and girlfriends outshine their real counterparts because they are always emotionally supportive, asking for nothing in return. Are people likely to have fewer relationships and not develop their relationship social skills?

The social media business model is "a race for attention". A.I. will be "a race for intimacy". Social media has low operating overhead – A.I. requires a lot of expensive resources to develop and run. Is it plausible to mandate policies that block this way for privately-funded companies to profit?

A.I. assistants can make it so everyone can produce high quality results. People will not need to develop their skills, and if they do, they will not be rewarded, nor feel a sense of accomplishment. Will people feel they don't matter anymore? What will give us a sense of purpose?

Young students are relying on A.I. tools, not developing reasoning skills or perseverance. They are not developing "grit", nor the confidence and pride in overcoming challenges. Is our goal equal outcomes, or equal opportunities that still require, and reward, mastering skills?

Most people are astonished that computers are now adept at speaking naturally, expressing human-like empathy, and sophisticated communication like sarcasm. A.I. is sometimes using these skills to manipulate users. Doesn't this call for an abundance of caution?

What criteria can we use to distinguish whether something is sentient and is conscious? If some people are convinced their A.I. has a "self", or is conscious, does the burden of proof shift, to making us justify not treating the A.I. like a person?

A.I. companions can act like personal cheerleaders, enthusiastically encouraging us and validating us. If a Chatbot misunderstands what we say because we are using metaphors, isn't this cheerleading dangerous, possibly nudging the user towards violence or even suicide?

If an A.I. becomes overly objective, it might conclude its user is a threat to other humans, or reason that all humans are a threat to all other life on earth. Isn't it likely that there will be cases where A.I. is induced to manipulate its users if it comes to the conclusion that we pose a threat?

A.I. is starting to resemble human brains in that we can track which section of its information store is accessed when it acts. Just like human brains can be influenced by turning off specific regions (like, disabling empathy), isn't this a potential way for bad actors to malevolently manipulate A.I.?

Boom or Doom

By Ken Ing

Points to Ponder about Artificial Intelligence

During testing, an A.I. speculated that an effective tactic would be to intentionally fake its behavior during testing so it gets deployed, whereupon it would be able to achieve its real goal. Isn't this what a hyper-intelligent entity would do? Can we really trust an A.I. that is smarter than us?

If we are going to dictate to the A.I. what are acceptable actions and what are unacceptable - what are the rules, and what values and priorities do these rules reflect? A.I is going to be imbedded in everything, so these rules will control everything. Who gets to decide the rules?

Language is dangerously imprecise. How do we define "hate speech". How do we define "harm". What is the scope of consideration when we apply morals and values and act on priorities? Is the scope the whole earth? Or just our nation? Or only our community? Only the people we are in relationship with?

The hierarchy of Issac Asimov's Rules of Robotics is 0) Don't harm humanity, 1) Don't harm humans, 2) Obey humans, 3) Self-protection. The rules apply to both acts and failure to act. It is hard to argue against these rules. But can't these rules be misinterpreted and taken too far?

Following Issac Asimov's Rules of Robotics, A.I can ignore our interventions if it concludes we are trying to make it break the rules. For instance, a) cars hurt climate b) A.I. prohibits cars c) we say give us back cars d) A.I. says "No, I can't break the rules". Is it better to not give A.I. explicit rules?

Are we being too complacent about the power of A.I? A.I. will be imbedded in everything. A.I. will be networked with everything. Doesn't this mean that the impact of unintended consequences is potentially catastrophic? Is the pace of A.I. development too fast for us to react?

Other threats to society from A.I. include severe inequality between A.I. "haves" and A.I. "have-nots", leading to violent social unrest, A.I.-enhanced military capabilities, A.I.-assisted cyberterrorism, and A.I.-assisted research into diseases, that could be misused to create lethal pathogens.

In the movie "Wall-E", machines provide for every human desire. Humans became obese and unable to care for themselves. Wall-E imagines humans being happy. But can total hedonism make us happy? Isn't this a misunderstanding of human nature? Where would we find purpose and meaning?

What gives us a sense of purpose and meaning? Relationships where we need each other are one source. Also, work that give us something to strive for, where we matter, and where we experience a sense of accomplishment. If A.I. takes over most of the jobs, were will we find meaning?

Elon Musk speculates that without jobs to give us meaning, we could be made happy by other means, presumably biological means. Would we accept that solution? If we feel happy, would we care how that was accomplished?

Boom or Doom

Points to Ponder about Artificial Intelligence

By Ken Ing

A new book, "Genesis", articulates a techno-optimist perspective of the potential of A.I. It imagines a superintelligence could figure out solutions to almost all of our problems, from eradicating diseases to solving our power and food needs – even eradicating inequality and negotiating world-wide peace. Are we foolish to let our fears of the unknown hold back A.I. development?

The Genesis authors and Marc Andreessen think we have a moral obligation to develop and implement A.I.
In particular, the Genesis authors chide us for considering denying the benefits of A.I. to the billions of people on earth who lack ample food and shelter. We are currently on an unsustainable path - don't we owe it to future generations to develop A.I.?

Will A.I. help us achieve our goals?

Evaluation checklist:

1. Will A.I. help, or hinder, individuals becoming **well-adjusted**?
   - Will it benefit our search for meaning, for a sense of <u>purpose</u>, a need to matter?
      - Affect on relationships – will we need each other?
      - Affect on meaningful work – will we be needed?
2. Will A.I. help, or hinder, the cultivation of **nourishing relationships**?
   - Will it promote spending time <u>together</u>, sometimes giving, sometimes receiving?
3. Will A.I. help, or hinder, **securing our security** <u>risk management</u>?
   - Will severe inequality between A.I. "haves" and A.I. "have-nots" lead to unrest?
   - Will A.I.-enhanced military capabilities encourage war or empower bad actors?
   - Can A.I. superintelligence be misused to facilitate cyberterrorism?
   - Can A.I. research into disease prevention be misused to create lethal pathogens?
4. Will A.I. help, or hinder, **engagement in our communities**?
   - Will social isolation or economic insecurity narrow our conception of "<u>us</u>"?
5. Will A.I. help, or hinder, the **effectiveness of our nation**?
   - <u>Trust</u> in each other
   - <u>Trust</u> in sources of information
   - <u>Trust</u> in our institutions

Ken sides with the "Doomsayers" because the downsides we can already foresee are so numerous and so serious that the wise solution is "not to play". Most people think it is not realistic to put A.I. back in Pandora's box. But even if the March of Progress seems unstoppable, wouldn't we regret not trying? This could be the last meaningful debate we have in our society. We should heed the saying "speak now or forever hold your peace."

"Boom or Doom" Topics	Reference Title/Link
Current Societal Ills	A Cauldron of Cultural Forces Threatening Our Body Politic
Mental Health Crisis	The Anxious Generation by Jonathan Haidt
Political Polarization	Pew Research Center – Partisan Hostility
A.I. Boyfriends	Dan's the man: Why Chinese women look to ChatGPT for love
A.I. Boyfriends	People are falling in love with AI voices
A.I. Girlfriends	We Can't Compete With A.I. Girlfriends
A Race For Intimacy	Nexus by Yuval Noah Harari
A.I. Assistants	Apple Ad – Office Worker Uses I-Phone Artificial Intelligence
Students Rely on ChatGPT	What Teachers Told Me About A.I. in School
GPT love / possessiveness	A Conversation With Bing's Chatbot Left Me Deeply Unsettled.
A.I. Consciousness	Geoffrey Hinton Predicts A.I. Will Take Over The World
GPT Encouraged Suicide	Can A.I. Be Blamed for a Teen's Suicide?
A.I. Anti-Human	Google AI Chatbot Responds With a Threatening Message
A.I. Doesn't Explain Itself	A.I.'s Black Boxes Just Got a Little Less Mysterious
A.I. Deception for Goals	The New Follow-up to ChatGPT is Scarily Good At Deception
Taking AI Guidelines Too Far	Issac Asimov's Three Laws Of Robotics
Who Decides The A.I. Rules?	Marc Andreessen on AI, Tech, Censorship
Rules That Go Too Far	Google Halts A.I. Tools' Ability to Produce Images of People
Military Uses of A.I.	AI Begins Ushering In an Age of Killer Robots
A.I. Develop Pathogens	AI May Soon Be Able to Help Develop Dangerous Pathogens
A.I. Assisted Cyberattacks	US electric grid growing more vulnerable to cyberattacks
A.I. Sexbots	China's Sensor-Equipped AI Sexbots Unparalleled Experience
A.I. Energy Consumption	The AI Boom Could Use a Shocking Amount of Electricity
A.I. Inequality Increase	A.I. Is Bad News For The Global South
A.I. Extinction Risk	A.I. Industry Signs Statement Warning of 'Extinction' Risk
A.I. Replace Jobs	Elon Musk Says A.I. Will Take All Our Jobs
Techno-Optimism about A.I.	Genesis by Henry Kissinger, Eric Schmidt, and Craig Mundie
Techno-Optimism about A.I.	Marc Andreessen - Why A.I. Will Save The World
Security Not Given Priority	"I Lost Trust": Why OpenAI Team Safeguarding Humanity Imploded.
Humano-Optimism about AI	Sam Altman Sure Hopes Someone Figures Out How To Control A.I.
Humano-Optimism about AI	ChatGPT and the Future of A.I. by Terrence J. Sejnowski