# PauseAI Montreal

(Questions at the end)

# Want

- Families, Loved ones
- Growth, Learning and self-improvement
- Creativity, Arts
- Safety for everyone
- Health
- Global cooperation



pauseai.ca/mtl/intro-slides
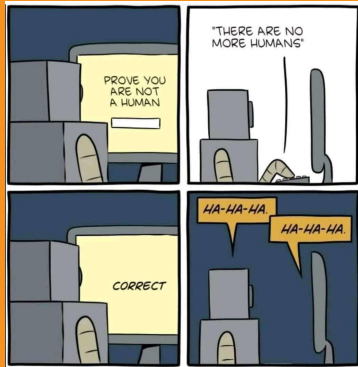
Humanity still in power

pauseai.ca/mtl/intro-slides

# AI Existential Risk

https://intelligence.org/the-problem/

1. We are on track to create artificial intelligence smarter than humans
2. It will have different goals
3. It will prevent us from reaching our goals for instrumental reasons
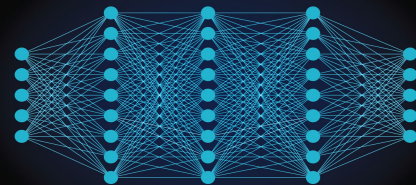






pauseai.ca/mtl/intro-slides

# AI is a Blackbox

Unlike other technologies, it is **grown** from data, not *programmed* with code, so even its designers don't know what it can do or why it makes one choice instead of another. Modern AI systems are hundreds of billions to trillions of numbers that no one understands.

```
 3    0.2232, 0.3242, 0.4252, 0.5262,
 4    0.6272, 0.7282, 0.8292, 0.9303,
 5    0.1313, 0.2323, 0.3333, 0.4343,
 6    0.5353, 0.6363, 0.7373, 0.8383,
 7    0.9393, 0.1404, 0.2414, 0.3424,
 8    0.4434, 0.5444, 0.6454, 0.7464,
 9    0.8474, 0.9484, 0.1495, 0.2505,
10    0.3515, 0.4525, 0.5535, 0.6545,
11    0.7555, 0.8565, 0.9575, 0.1586,
12    0.2596, 0.3606, 0.4616, 0.5626,
13    0.2232, 0.3242, 0.4252, 0.5262,
14    0.6272, 0.7282, 0.8292, 0.9303,
15    0.1313, 0.2323, 0.3333, 0.4343,
16    0.5353, 0.6363, 0.7373, 0.8383,
17    0.9393, 0.1404, 0.2414, 0.3424,
18    0.4434, 0.5444, 0.6454, 0.7464,
19    0.8474, 0.9484, 0.1495, 0.2505,
20    0.3515, 0.4525, 0.5535, 0.6545,
21    0.7555, 0.8565, 0.9575, 0.1586,
22    0.2596, 0.3606, 0.4616, 0.5626,
23    0.2232, 0.3242, 0.4252, 0.5262,
24    0.6272, 0.7282, 0.8292, 0.9303,
25    0.1313, 0.2323, 0.3333, 0.4343,
26    0.5353, 0.6363, 0.7373, 0.8383,
27    0.9393, 0.1404, 0.2414, 0.3424,
28    0.4434, 0.5444, 0.6454, 0.7464,
29    0.8474, 0.9484, 0.1495, 0.2505,
30    0.3515, 0.4525, 0.5535, 0.6545,
31    0.7555, 0.8565, 0.9575, 0.1586,
32    0.2596, 0.3606, 0.4616, 0.5626
```

weights/numbers -> (LLM)

<- instructions (typical program)

```
32
33    if (pos.castling_impeded(Cr) || !pos.can_castle(Cr))
34        return moveList;
35
36    // After castling, the rook and king final positions are the same in Chess960
37    // as they would be in standard chess.
38    Square kfrom = pos.square<KING>(us);
39    Square rfrom = pos.castling_rook_square(Cr);
40    Square kto = relative_square(us, KingSide ? SQ_G1 : SQ_C1);
41    Bitboard enemies = pos.pieces(~us);
42
43    assert(!pos.checkers());
44
45    const Square K = Chess960 ? kto > kfrom ? DELTA_W : DELTA_E
46                              : KingSide   ? DELTA_W : DELTA_E;
47
48    for (Square s = kto; s != kfrom; s += K)
49        if (pos.attackers_to(s) & enemies)
50            return moveList;
51
52    // Because we generate only legal castling moves we need to verify that
53    // when moving the castling rook we do not discover some hidden checker.
54    // For instance an enemy queen in SQ_A1 when castling rook is in SQ_B1.
55    if (Chess960 && (attacks_bb<ROOK>(kto, pos.pieces() ^ rfrom) & pos.pieces(~us,
56        return moveList;
57
58    Move m = make<CASTLING>(kfrom, rfrom);
59
60    if (Checks && !pos.gives_check(m, *ci))
61        return moveList;
62    else
63        (void)ci; // Silence a warning under MSVC
64
65    *moveList++ = m;
66    return moveList;
67  }
```

## PAUSE AI

pauseai.ca/mtl/intro-slides

# Prominent people on AI Existential Risk

https://pauseai.info/xrisk

**Stuart Russell**, writer of the #1 textbook on Artificial Intelligence used in most AI studies: "If we pursue [our current approach], then we will eventually lose control over the machines"

**Yoshua Bengio**, deep learning pioneer and winner of the Turing Award: "… rogue AI may be dangerous for the whole of humanity […] banning powerful AI systems (say beyond the abilities of GPT-4) that are given autonomy and agency would be a good start"

**Geoffrey Hinton**, the "Godfather of AI" and Turing Award winner, left Google to warn people of AI: "This is an existential risk"

**Sam Altman** (yes, the CEO of OpenAI who builds ChatGPT): "Development of superhuman machine intelligence is probably the greatest threat to the continued existence of humanity."

.

**II PAUSE AI**

pauseai.ca/mtl/intro-slides

# Other AI Risks

https://pauseai.info/risks

- Present dangers
    - Fake news, polarization and threatening democracy
    - Deepfakes-powered harassment and scams
    - Biases and discrimination
    - Job loss, economic inequality and instability
    - Mental health, addiction and disconnection between people
    - Automated investigation (loss of privacy)
    - Environmental risks
    - Autonomous weapons
- Near future dangers
    - Power accumulation and tyranny
    - Biological weapons
    - Computer viruses and cybersecurity attacks
    - Existential Risk
    - Human disempowerment
    - Digital sentience
    - Suffering risks

pauseai.ca/mtl/intro-slides

# PauseAI Global Treaty

**Implement a temporary pause on the training of the most powerful general AI systems**, until we know how to build them safely and keep them under democratic control.

II PAUSE AI

pauseai.ca/mtl/intro-slides

# PauseAI Global Theory of change

https://pauseai.info/theory-of-change

Grow the movement -> Everyone email MPs -> AIS Discussion -> AI law / Treaty

1. **Grow the movement**. The larger our group is, the more we can do. [...] online community building, and fostering local communities. Read more about our growth strategy [...].
2. **Protests**. Protests [...] public awareness [...]. Because our subject is relatively new, even small protests can get very good media coverage.
3. **Lobbying**. Every volunteer can become an amateur lobbyist. We send emails to politicians, we meet with them, [...].
4. **Inform the public**. We make people aware of the risks we're facing and what we can do to prevent them. [...] publishing articles, videos, images and posts on social media. We join podcasts, give talks, and organize events. Read about our communication strategy.

**⏸ PAUSE AI**

pauseai.ca/mtl/intro-slides

# What we do at PauseAI Montreal

- Find other concerned people and Informs the public
  - Flyering
  - Put up posters around the city
  - Find people already concerned about AI at meetups around montreal
- Talk to politicians and policy-makers
  - Email and talk to local politicians to push them to vote for AI regulation
- Organizes events
  - Social Meetups
- Keep ourselves informed
  - Book club: *If Anyone Builds It, Everyone Dies*
  - Meet and discuss how AI should be regulated, deployed, to make it safe and help humanity rather than cause issues

Links

- https://intelligence.org/the-problem/
- https://pauseai.info/xrisk
- https://pauseai.info/risks
- https://pauseai.info/proposal
- https://pauseai.info/theory-of-change

Questions? Questions!

pauseai.ca/mtl/intro-slides