

# Why AI Companies Will Always Choose Profit Over Your Safety

Date: April 30, 2026 | Model: anthropic-batch:claude-opus-4-7

Source: Screenshot (OCR via AI)

---

## Contents

1. Explanation (Ages 14–18)
2. Key Terms Glossary
3. Reading Comprehension Quiz (10 questions)
4. Answer Key with Explanations

*Note: the original article is provided as a separate file (attached to the email or downloadable from the website).*

## 1. Explanation (Ages 14–18)

---

*AI executives love talking about saving humanity from their own creations. But when profit collides with principle, capitalism has a 200-year track record – and humanity's safety usually loses.*

### What's Going On?

Financial Times columnist Robert Armstrong argues that despite AI leaders' lofty promises about ethical guidelines and safety, AI companies will ultimately behave like any other corporation: they'll maximise shareholder returns within whatever the law allows. When profit conflicts with internal safety principles, he writes, profit will win every time.

He points to staggering numbers – Big Tech 'hyperscalers' plan to invest over \$600bn this year, and OpenAI alone raised \$122bn in a single funding round. Investors expect aggressive growth, and CEOs who slow down for safety reasons risk being fired. Armstrong concludes that meaningful AI safety must come from smart regulation, not corporate self-restraint.

### How To Think About It

The core idea here is that incentive structures beat good intentions almost every time. Don't focus on what executives say in interviews – look at who pays them and what those people want.

- Think of social media circa 2010: Facebook executives genuinely believed they were 'connecting the world.' But ad-revenue incentives pushed engagement algorithms toward outrage and addiction. Good intentions, bad incentives, predictable outcome.
- Or think of the 2008 financial crisis: bank CEOs knew mortgage-backed securities were risky, but the bonus structures rewarded short-term deal volume. Individual ethics couldn't override systemic incentives.

### Key Things To Know

- AI startups raised \$73bn in just the first quarter of 2025 – the capital comes from investors demanding aggressive returns.
- Anthropic CEO Dario Amodei and OpenAI publish 'safety guidelines' (like OpenAI's 'Model Spec'), but these are voluntary and have no legal force.
- When OpenAI reportedly missed user targets last week, the Nasdaq dropped – showing how investor pressure punishes any hint of slower growth.
- Armstrong's non-obvious point: the profit motive normally serves society well by encouraging innovation. The problem is specific to industries where externalities (harms to third parties) are massive.
- What people get wrong: assuming CEOs who sound sincere about safety can actually act on it. Their job is to satisfy shareholders; safety talk that hurts revenue gets them replaced.

### Why It Matters

You're going to live and work in an AI-shaped economy – applying to college through AI-screened

systems, competing for jobs against AI tools, and possibly building careers in fields that don't exist yet. How (or whether) governments regulate AI in the next five years will shape the labour market you enter, your privacy, and even what counts as 'truth' online. This isn't an abstract policy debate; it's the operating system of your adult life.

### **The Bigger Picture**

Armstrong proposes treating AI more like explosives – emphasising liability for harms rather than relying on a single sweeping law. Watch for fights over targeted rules: copyright, deepfakes, AI in hiring, autonomous weapons. The historical parallel is the early 20th century, when industries from cars to pharmaceuticals went from unregulated free-for-alls to carefully governed sectors – usually only after disasters forced governments to act. The question is whether AI regulation arrives before or after its equivalent catastrophe.

## 2. Key Terms Glossary

---

### **Hyperscalers**

The handful of massive cloud-computing companies (Microsoft, Google, Amazon, Meta) with the data-centre infrastructure to train and run frontier AI models.

### **Externalities**

Costs or benefits a business activity creates for third parties who aren't part of the transaction – for example, pollution from a factory affects neighbours who never bought its products.

### **Shareholder returns**

The profit a company generates for the people who own its stock, typically through rising share prices and dividends. Public-company executives are legally and culturally expected to prioritise this.

### **Model Spec**

OpenAI's published document describing how its AI models are supposed to behave – essentially a voluntary code of conduct, not a binding regulation.

### **Agency law**

The branch of law governing situations where one party (an 'agent') acts on behalf of another. Armstrong suggests applying it to AI systems acting on behalf of humans.

### **Monolithic law**

A single, sweeping piece of legislation that tries to cover an entire complex topic at once – as opposed to many narrow, targeted rules.

### **Clickbait**

Online content designed to provoke clicks through sensational or misleading framing rather than substance. OpenAI used the term to dismiss a Wall Street Journal report.

### 3. Reading Comprehension Quiz

---

Circle the best answer for each question.

**Q1.** Which choice best states the central idea of the passage?

- A) AI technology will eliminate most human jobs within a generation, much as cars replaced horses.
- B) AI companies, despite safety rhetoric, will prioritise profit, so meaningful safeguards must come through regulation.
- C) Government regulation of emerging technologies always produces unintended consequences and should be avoided.
- D) AI executives like Sam Altman and Dario Amodei are insincere about their concerns regarding AI risks.

**Q2.** According to the passage, what happens when a company's internal ethical principles conflict with its profit motive?

- A) Companies typically pause operations until the conflict is fully resolved internally.
- B) Regulators step in to enforce the company's stated ethical principles directly.
- C) The profit motive prevails, since shareholder returns are the overriding obligation.
- D) Executives publicly resign rather than violate their stated ethical commitments.

**Q3.** The passage indicates that the horse-and-automobile analogy is used by AI optimists primarily to suggest that:

- A) Workers displaced by AI will eventually find better lives, as horses moved from labour to leisure.
- B) AI will follow the same gradual adoption curve as automobiles did during the early 1900s.
- C) Governments should intervene early to protect workers from technological displacement.
- D) Technological change inevitably creates more jobs than it destroys in the long run.

**Q4.** As used in the passage, the word 'externalities' most nearly means:

- A) external advertising and marketing strategies aimed at consumers.
- B) side effects of business activity that affect parties outside the transaction.
- C) international operations a company conducts beyond its home country.
- D) outside investors who fund a company's growth and expansion.

**Q5.** As used in the passage, the word 'noise' most nearly means:

- A) loud commotion that interrupts business proceedings.
- B) background interference in a digital communication system.
- C) talk that doesn't reflect the speaker's actual incentives.
- D) public criticism aimed at influencing corporate behaviour.

**Q6.** Which statement about AI safety guidelines can most reasonably be inferred from the passage?

- A) They will become legally binding once regulators formally endorse them.
- B) They are likely to be overridden whenever they threaten company revenues.
- C) They have already prevented several major AI-related public harms.
- D) They are written primarily by government officials rather than companies.

- Q7.** The passage suggests that the example of OpenAI's response to the Wall Street Journal report primarily illustrates:
- A) the media's tendency to exaggerate problems at successful technology companies.
  - B) the extreme sensitivity of AI firms to anything threatening revenue expectations.
  - C) the legal obligations companies face when responding to journalistic inquiries.
  - D) the difficulty of accurately measuring user growth at fast-moving startups.
- Q8.** The author's tone throughout the passage is best described as:
- A) alarmed and urgent, warning of imminent civilizational collapse from AI.
  - B) sceptical and pragmatic, accepting capitalism while questioning self-regulation.
  - C) celebratory and optimistic about AI's potential to transform human labour.
  - D) nostalgic and resistant, longing for an era before disruptive technologies.
- Q9.** Based on the passage, what kind of regulation does the author suggest would be most effective for AI?
- A) A single comprehensive law that protects specific job categories from automation.
  - B) Targeted rules addressing specific harms, with strong liability for damages caused.
  - C) Voluntary industry guidelines reinforced by public pressure and media scrutiny.
  - D) International treaties that uniformly govern all AI development worldwide.
- Q10.** Which choice provides the best evidence for the answer to the previous question?
- A) 'A lot of people once drove horse-drawn carts and made buggy whips, they say.'
  - B) 'It should match specific regulatory tools to specific harms... rather than taking the form of a monolithic law.'
  - C) 'Everyone in the AI industry acknowledges this. It is expressed in OpenAI's Model Spec guidelines.'
  - D) 'The amounts of money AI has attracted are staggering... \$600bn in the space this year alone.'

**My Score:** \_\_\_\_\_ / 10

## 4. Answer Key with Explanations

---

**Q1.** Which choice best states the central idea of the passage?

**Answer: B**

The passage builds toward the conclusion that capitalism's logic forces profit-maximisation, so safety requires external rules – Armstrong explicitly calls for regulation in the final paragraphs. Option D is the main trap (Trap C: plausible real-world claim) but Armstrong actually says he believes Amodei is sincere; the problem is structural, not personal. SAT Tip: when asked for the central idea, pick the option that captures the passage's overall argument, not a single sub-claim – even a true sub-claim is wrong if it's narrower than the main thesis.

**Q2.** According to the passage, what happens when a company's internal ethical principles conflict with its profit motive?

**Answer: C**

Armstrong writes plainly that 'when the law of profit conflicts with the company's internal principles, profit will win every time.' Option D (Trap C) sounds noble and occasionally happens in real life, but the passage makes the opposite claim. SAT Tip: when a question asks 'according to the passage,' your answer must be directly supported by the text – even widely-true real-world facts are wrong if the passage doesn't say them.

**Q3.** The passage indicates that the horse-and-automobile analogy is used by AI optimists primarily to suggest that:

**Answer: A**

The passage states that optimists argue 'horses' lives got better as they went from work animals to luxury items.' Option D (Trap C) is a real claim economists often make about technology – but it's not the specific claim Armstrong attributes to the optimists in this passage. SAT Tip: watch for answer choices that swap the passage's actual analogy for a different (even related) economic argument – stay literal.

**Q4.** As used in the passage, the word 'externalities' most nearly means:

**Answer: B**

Armstrong refers to 'economic externalities they create when they are forced to' pay for them – meaning effects on parties outside the company that companies normally ignore. Option A (Trap B) borrows the 'external' root word but misuses it. SAT Tip: on vocab-in-context, substitute each option for the word in the original sentence – only the correct meaning will preserve what the sentence is actually saying.

**Q5.** As used in the passage, the word 'noise' most nearly means:

**Answer: C**

Armstrong writes that Amodei's words 'are just noise' because 'the relevant incentive structures don't care' – meaning the words have no real effect on outcomes. Option A (Trap B) takes the everyday meaning of noise but doesn't fit the metaphor. SAT Tip: vocab-in-context questions almost always test a word's secondary or metaphorical meaning, not its first dictionary definition – if the obvious answer feels too easy, it's probably the trap.

**Q6.** Which statement about AI safety guidelines can most reasonably be inferred from the passage?

**Answer: B**

The passage's whole argument is that internal principles lose to profit pressure, and notes how investors punished OpenAI for missing growth targets – implying safety commitments would similarly buckle under such pressure. Option A (Trap C) sounds plausible as a future scenario but isn't suggested by the passage. SAT Tip: inference

questions reward conclusions that follow directly from the passage's logic, not predictions that merely sound reasonable.

**Q7.** The passage suggests that the example of OpenAI's response to the Wall Street Journal report primarily illustrates:

**Answer: B**

Armstrong introduces the anecdote to show that 'the industry's sensitivity to revenue growth expectations is extreme' – the Nasdaq dropped, and OpenAI scrambled to call the story 'clickbait.' Option A (Trap A: right scope, wrong direction) flips the point – Armstrong uses the example to criticise the company's reaction, not the media. SAT Tip: when an example is given, ask 'what claim is this example meant to support?' rather than guessing from the example alone.

**Q8.** The author's tone throughout the passage is best described as:

**Answer: B**

Armstrong calls capitalism 'beautiful' for encouraging innovation but doubts companies will police themselves, calling executive safety talk 'noise.' Option A (Trap B) uses passage-adjacent language ('warn,' 'fears') but overstates the alarm – Armstrong is measured, not panicked. SAT Tip: identify tone by the author's adjectives and verbs, not the seriousness of the topic – a writer can discuss existential risks in a calm, analytical voice.

**Q9.** Based on the passage, what kind of regulation does the author suggest would be most effective for AI?

**Answer: B**

Armstrong explicitly rejects 'a monolithic law' and instead calls for 'specific regulatory tools to specific harms' and for taking 'liability' seriously. Option A (Trap B) recycles passage vocabulary ('specific') but combines it incorrectly – Armstrong opposes protecting job categories. SAT Tip: when an author offers a list of recommendations, check whether each answer choice combines those elements correctly – distractors often mix real ingredients in wrong proportions.

**Q10.** Which choice provides the best evidence for the answer to the previous question?

**Answer: B**

This quote directly states Armstrong's preferred regulatory approach – targeted rules rather than one sweeping law – which matches the previous answer. Option C (Trap B) is from the passage and discusses guidelines, but it describes industry self-regulation, which Armstrong is criticising, not endorsing. SAT Tip: on evidence-pairing questions, find the line that supports your previous answer FIRST, then match it to an option – never pick evidence that merely 'sounds related' to the topic.