

GRADES 9–12 · AGES 14–18 · 60 MINUTES · POLICY DEBATE

Who owns what the model learned?

THE PROPOSITION

"Training an AI model on copyrighted work without permission should require a license."

A policy debate on training data and copyright, with the affirmative proposing a licensing plan and the negative attacking it or offering a counter-plan. Students meet fair use, the open-weights question, and the practical problem that the training already happened.

STUDENTS WILL BE ABLE TO

- Explain what training data is and why copyright law did not anticipate it.
- Propose or attack a concrete policy with attention to enforcement, cost, and who benefits.
- Evaluate the fair-use argument on its merits rather than on who is making it.

VOCABULARY FROM THE FIELD GUIDE

Training data · Copyright & training · Consent (data use) · Open weights / open models · Closed / proprietary model · Moat (competitive advantage)

WHAT'S IN THIS PACK

- Teacher guide: background, vocabulary, agenda, format, positions, questions, roles, rubric.
- Student handouts (photocopy one per student): position organizer, sentence starters, exit ticket.
- Standards: CCSS ELA SL.9–10.1 / SL.11–12.1; CCSS ELA W.9–10.1 / W.11–12.1 (argument with counterclaims); C3 Framework D2.Civ.13 (public policy analysis); ISTE Students 2.c (intellectual property).

PRINT

US Letter, plain paper, duplex. Teacher pages once; student pages one per student. Free to photocopy for classroom and homeschool use. Not for resale.

Background for the teacher

Read this before class. Students get a version of the same facts on their handout; you get the framing and the traps.

Modern models are trained on text and images scraped from the web at a scale that includes most published books, news archives, code repositories, and art portfolios – generally without permission or payment. Authors, artists, and publishers have sued. Developers argue the training is 'fair use': transformative, non-expressive, more like reading than copying. Courts are split and the question is unresolved.

The policy options are roughly: (a) training is fair use, no license needed; (b) opt-out – creators can block training but must act; (c) opt-in licensing – no training without permission, likely through collective licenses like music radio uses; (d) compulsory license – anyone may train, but must pay a set rate into a fund. Each shifts money and power differently.

Two complications students should reach: a licensing regime favors the largest companies who can afford it and entrenches their moat, while penalizing open-weights and academic models; and the frontier models were already trained, so any rule is partly about the future and partly about restitution.

RUNNING DISCUSSION IN GRADES 9–12

Stakes are real: college, jobs, creative work, civil rights. Students can steelman, cross-examine, and write policy.

- Insist on the steelman. Before rebutting, a speaker restates the other side's best version until they say 'yes, that's it.'
- Cross-examination is a question period, not a speech. Two minutes, questions only, no editorializing.
- Push past 'it depends.' Ask: depends on what, exactly – and who decides?
- Close with a written position. Speaking is the practice; the paragraph is the assessment.

Vocabulary

Pulled from the Hi, Bot AI concepts field guide (hibot.space/learn/concepts). Read each card aloud, then have the class rewrite it in ten words or fewer.

Training data

The examples a model learned from – often the whole internet, licensed corpora, or both.

Training data is the text, images, or logs used during training. Quality, consent, copyright, and representation in that data shape model behavior. When people ask "where did it learn that?", they're asking about training data and fine-tuning – usually opaque for closed models.

WHERE PEOPLE ARGUE

Copyright, consent, and "fair use for training" are active legal and political fights worldwide.

Copyright & training

Who owns the inputs models learned from – and who owns the outputs.

Copyright debates ask whether scraping books, art, or code to train models infringes creators' rights, and whether AI-generated works are protectable. Laws differ by country and are still unsettled. For users: assuming "the model said it, so I can publish it" is risky – verify rights on sources and outputs.

Consent (data use)

Whether people agreed to their data being used to train or fine-tune systems.

Consent questions arise when personal writing, medical records, student work, or biometric data might enter a training or fine-tuning pipeline. Schools and companies often ban pasting identifiable student data into public models for this reason. "Public on the web" is not the same as "consented for AI training" in many ethical frameworks.

Open weights / open models

Models whose trained parameters are published for anyone to download and run.

Open-weight models release the actual trained network (often with a license governing use). That enables local running, inspection, and fine-tuning without calling a vendor API. "Open" varies – some licenses restrict commercial use or require sharing derivatives. Open weights are not the same as open training data.

WHERE PEOPLE ARGUE

Security vs. transparency: open models aid research and self-hosting; critics worry about misuse without gatekeeping.

Closed / proprietary model

A model you access through an API – weights not public.

Closed models run on the provider's infrastructure; you send prompts and receive outputs under their terms. You trade control and inspectability for convenience, safety filtering, and scale. Most consumer chatbots are closed models behind accounts and rate limits.

Moat (competitive advantage)

Why one AI product keeps customers when models commoditize.

In business debates, a moat is defensibility: proprietary data, distribution, brand, regulatory approval, workflow lock-in, or fine-tuned models on exclusive corpora. Base models trend toward commodity; applications fight on trust, integration, and outcomes.

Warm-up

6 minutes.

Ask: 'You read a hundred novels and then write one. Do you owe the hundred authors anything?' Then: 'A company reads a hundred million and sells a product that writes novels. Same answer?' Pair-share what changed – scale, purpose, or who profits.

Agenda - 60 minutes

Minutes	Segment	What happens
0-6	Warm-up	Human reader vs. model trainer. What changed?
6-18	Background + policy menu	Read the sheet. Teams map the four policy options on a grid: who pays, who benefits, who can enforce.
18-28	Case construction	Affirmative builds a plan (which policy, who administers, rate, exemptions). Negative prepares attacks and an optional counter-plan.

Minutes	Segment	What happens
28–52	Policy debate	Aff constructive 3 · Neg cross-ex 2 · Neg constructive 3 · Aff cross-ex 2 · Neg rebuttal 2 · Aff rebuttal 2. Judges score plan, evidence, clash.
52–60	Written position	Exit ticket: which policy, why, and who loses under it.

How policy debate runs

The proposition is a concrete policy ("Our school should...", "The law should..."). Affirmative proposes a plan and defends it; Negative attacks the plan and may offer a counter-plan. Constructive speeches (3 min), cross-examination (2 min), rebuttals (2 min). Judges score on plan, evidence, and clash.

GROUND RULES – POST THESE

1. Attack ideas, never people. "That claim has a hole" – not "you're wrong."
2. One voice at a time. The listener's job is to be able to repeat what was said.
3. Evidence beats volume. A quiet fact outranks a loud opinion.
4. Changing your mind is winning. Say so out loud when it happens.
5. Plans must name who administers and who pays.
6. Cross-examination is questions only.

The two positions

Starter arguments are scaffolds, not scripts. Strong teams find their own. Assign sides where the format calls for it; students record their real view on the exit ticket.

Affirmative – require a license

Creators' work has value that trained a commercial product; the law should require permission and payment, as it does for music and film.

- A model is not a reader. It's a product built from the work at industrial scale.
- Collective licensing already works for radio, streaming, and photocopying.
- Without a license, the incentive to create the next generation of training data disappears.
- Opt-out puts the burden on the person harmed, which is backwards.

Negative – training is fair use, or the plan fails

Learning patterns from public work is transformative and non-expressive; a licensing regime would entrench the richest firms and kill open models.

- The model doesn't store or reproduce the works; it learns statistics – closer to reading than copying.
- Only the largest companies can afford to license the internet. That's a moat, not justice.
- Open-weights and academic models can't pay; the plan hands AI to three firms.
- Counter-plan: an opt-out registry plus a compulsory license fund with a fixed rate – payment without gatekeeping.

Discussion questions

OPENING

1. Is a model 'reading' or 'copying'? What test would tell you?

CORE

1. Who benefits under each policy option – and who is locked out?
2. If a licensing rule makes open models impossible, is that an acceptable cost?
3. The models already exist. Does your plan handle restitution, or only the future?

CLOSING

1. Choose one policy. Name who loses under it and why that's acceptable.

Roles

Role	Job
Affirmative constructive	Presents the plan: policy, administration, rate, exemptions.
Negative constructive	Attacks the plan on enforcement, cost, and moat; may present a counter-plan.
Cross-examiner (each side)	Two minutes of questions only.
Rebuttal (each side)	Steelman, then answer; weigh impacts.
Judge panel	Scores on plan quality, evidence, and clash. Names the deciding issue.

Assessment rubric

Share this before the debate. Judges and self-assessors use the same sheet. 'Meets' is the target; note above or below in the margin.

Criterion	Meets
Argument	Claim, warrant, and evidence are explicit and connected.
Steelman	Presents the strongest version of the opposing view before attacking it.
Clash	Rebuttals respond to what was actually said, not a strawman.
Precision	Uses lesson vocabulary accurately and distinguishes is from ought.
Written position	Closing paragraph takes a defensible position and concedes one real cost.

Exit ticket prompts

1. The policy I'd choose: ____.
2. Who loses under it: _____. Why that's acceptable: _____.

3. The strongest argument against my choice: ____.

Extension

Read the summary of one active lawsuit over AI training data. Identify: the plaintiff, the fair-use argument, the plaintiff's answer, and what a win for each side would change.

SEND HOME

Ask at home: 'If something you wrote or made was used to train a model, would you want money, credit, a veto, or nothing? Why?'

STANDARDS ALIGNMENT

- CCSS ELA SL.9-10.1 / SL.11-12.1
- CCSS ELA W.9-10.1 / W.11-12.1 (argument with counterclaims)
- C3 Framework D2.Civ.13 (public policy analysis)
- ISTE Students 2.c (intellectual property)

More lessons by grade band, and the full AI concepts field guide, at hibot.space/learn/debates. Tell us how it went: hello@hibot.space.

Who owns what the model learned?

"Training an AI model on copyrighted work without permission should require a license."

WHAT YOU NEED TO KNOW

Modern models are trained on text and images scraped from the web at a scale that includes most published books, news archives, code repositories, and art portfolios – generally without permission or payment. Authors, artists, and publishers have sued. Developers argue the training is 'fair use': transformative, non-expressive, more like reading than copying. Courts are split and the question is unresolved.

The policy options are roughly: (a) training is fair use, no license needed; (b) opt-out – creators can block training but must act; (c) opt-in licensing – no training without permission, likely through collective licenses like music radio uses; (d) compulsory license – anyone may train, but must pay a set rate into a fund. Each shifts money and power differently.

Two complications students should reach: a licensing regime favors the largest companies who can afford it and entrenches their moat, while penalizing open-weights and academic models; and the frontier models were already trained, so any rule is partly about the future and partly about restitution.

WORDS WE'LL USE

Training data – The examples a model learned from – often the whole internet, licensed corpora, or both.

Copyright & training – Who owns the inputs models learned from – and who owns the outputs.

Consent (data use) – Whether people agreed to their data being used to train or fine-tune systems.

Open weights / open models – Models whose trained parameters are published for anyone to download and run.

Closed / proprietary model – A model you access through an API – weights not public.

Moat (competitive advantage) – Why one AI product keeps customers when models commoditize.

MARK IT UP

Underline one sentence you agree with. Circle one you doubt. Write one question in the margin. Bring all three to the discussion.

Position organizer

Proposition: "Training an AI model on copyrighted work without permission should require a license."

Name: _____ Side (assigned or chosen): _____

OUR CLAIM, IN ONE SENTENCE

THREE REASONS – AND THE EVIDENCE FOR EACH

Reason 1:

Evidence (a fact, a definition, or a specific example):

Reason 2:

Evidence (a fact, a definition, or a specific example):

Reason 3:

Evidence (a fact, a definition, or a specific example):

WHAT THE OTHER SIDE WILL SAY – AND OUR REPLY

Their best argument:

Our reply:

VOCABULARY I WILL USE CORRECTLY

☐ Training data ☐ Copyright & training ☐ Consent (data use) ☐ Open weights / open models ☐ Closed / proprietary model ☐ Moat (competitive advantage)

Sentence starters

Use these until you don't need them. The goal is to respond to what was actually said.

- Our plan is ____, administered by ____, funded by ____.
- Under your plan, who pays? Who's exempt?
- The fair-use argument holds only if ____.
- Our counter-plan solves ____ without ____.
- That's a cost we accept because ____.

GROUND RULES

1. Attack ideas, never people. "That claim has a hole" – not "you're wrong."
2. One voice at a time. The listener's job is to be able to repeat what was said.
3. Evidence beats volume. A quiet fact outranks a loud opinion.
4. Changing your mind is winning. Say so out loud when it happens.
5. Plans must name who administers and who pays.
6. Cross-examination is questions only.

HOW POLICY DEBATE WORKS

The proposition is a concrete policy ("Our school should...", "The law should..."). Affirmative proposes a plan and defends it; Negative attacks the plan and may offer a counter-plan. Constructive speeches (3 min), cross-examination (2 min), rebuttals (2 min). Judges score on plan, evidence, and clash.

YOUR ROLE TODAY

Role: _____

Role	Job
Affirmative constructive	Presents the plan: policy, administration, rate, exemptions.
Negative constructive	Attacks the plan on enforcement, cost, and moat; may present a counter-plan.
Cross-examiner (each side)	Two minutes of questions only.
Rebuttal (each side)	Steelman, then answer; weigh impacts.
Judge panel	Scores on plan quality, evidence, and clash. Names the deciding issue.

LISTENING TALLY

Keep count while others speak. Claims: ____ Evidence: ____ Questions: ____ Vocabulary used: ____
Minds changed: ____

Exit ticket

"Training an AI model on copyrighted work without permission should require a license."

Name: _____ Date: _____

The policy I'd choose: ____.

Who loses under it: _____. Why that's acceptable: _____.

The strongest argument against my choice: _____.

SELF-ASSESSMENT

Criterion	Meets looks like	Me today (above / meets / below)
Argument	Claim, warrant, and evidence are explicit and connected.	
Steelman	Presents the strongest version of the opposing view before attacking it.	
Clash	Rebuttals respond to what was actually said, not a strawman.	
Precision	Uses lesson vocabulary accurately and distinguishes is from ought.	
Written position	Closing paragraph takes a defensible position and concedes one real cost.	

TAKE IT HOME

Ask at home: 'If something you wrote or made was used to train a model, would you want money, credit, a veto, or nothing? Why?'