INVASION OF THE PLAUSIBLE SLOP

Invasion of the Plausible Slop

Generative AI and Open Source Cybersecurity

Dr. Kaylea Champion – kaylea@uw.edu
https://www.kayleachampion.com
https://communitydata.science
University of Washington | Bothell

FOSSY Science of Community—August 2, 2025

Community Data Science Collective

UNIVERSITY of WASHINGTON

My research interests

- I study how open source works and how to make it better

My research interests

- I study how open source works and how to make it better
- Including the risks we face and the ways we can prevail

My research interests

- I study how open source works and how to make it better
- Including the risks we face and the ways we can prevail
- But I'm not alone...


Unique
Knives
JAZZ

The key technologies behind generative AI are open source.

- Infrastructure and orchestration

The key technologies behind generative AI are open source.

- Infrastructure and orchestration
- Programming languages

The key technologies behind generative AI are open source.

- Infrastructure and orchestration
- Programming languages
- Training data

PyTorch

Learn Community Projects Docs Blog & News About JOIN

Get Started

Choose Your Path: Install PyTorch Locally or Launch Instantly on Supported Cloud Platforms

Get started →

Marketplace > Programming Languages > GitHub Copilot

GitHub Copilot

GitHub github.com | 44,764,208 installs | ★★★★

Your AI pair programmer

Installation

Launch VS Code Quick Open (Ctrl+P), paste the following command, and pres

```
ext install GitHub.copilot
```

Copy | More Info

Version History Q & A Rating & Review

Copilot - Your AI peer programmer

Copilot is an AI peer programming tool that helps you write code faster and smarter.

Copilot adapts to your unique needs allowing you to select the best model for your project, cus with custom instructions, and utilize agent mode for AI-powered, seamlessly integrated pe ng sessions.

GitHub Copilot Free!

learn Install User Guide API Examples Community More 1.7.1 (stable)

scikit-learn

Machine Learning in Python

Getting Started | Release Highlights for 1.7

- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Hugging Face Search models, datasets, users...

Main Tasks Libraries Languages Licenses Other

Tasks

Text Generation | Any-to-Any | Image-Text-to-Text
Image-to-Text | Image-to-Image | Text-to-Image
Text-to-Video | Text-to-Speech | + 42

Parameters

< 1B 6B 12B 32B 128B > 500B

Libraries

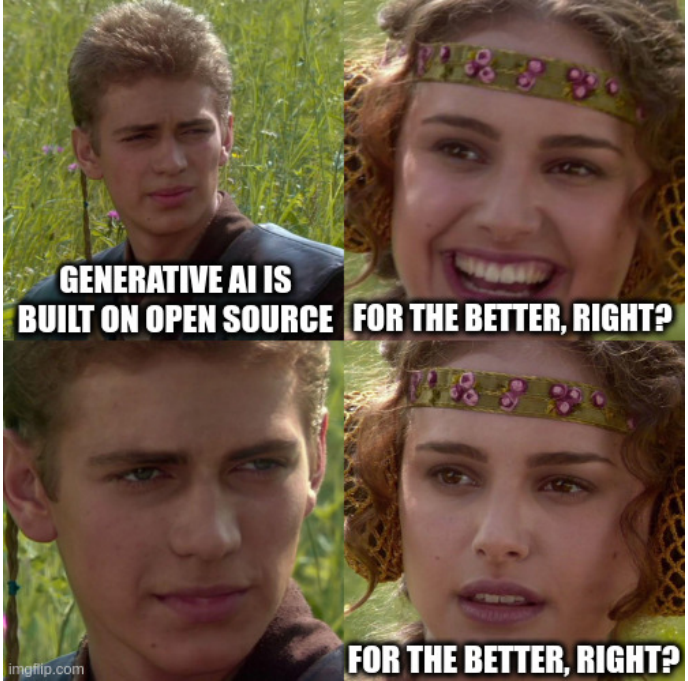
PyTorch | TensorFlow | JAX | Transformers

Wikipedia Search

Welcome to Wikipedia,

the free encyclopedia that anyone can edit.

107,250 active editors · 7,029,771 articles in English


GENERATIVE AI IS
BUILT ON OPEN SOURCE
FOR THE BETTER, RIGHT?
FOR THE BETTER, RIGHT?
imgflip.com

Cue the Slop

ars TECHNICA

SECTIONS FORUM SUBSCRIBE SIGN IN

AND MAKE IT SOUND ALARMING

Open source project curl is sick of users submitting "AI slop" vulnerabilities

"One way you can tell is it's always such a nice report," founder tells Ars.

KEVIN PURDY – UPDATED MAY 7, 2025 9:49 AM | 132

Credit: Aurich Lawson | Getty Images

"A threshold has been reached. We are effectively being DDoSed. If we could, we would charge them for this waste of our time," wrote Daniel Stenberg, original

HOME • BLOG • RSS • ABOUT

Don't bring slop to a slop fight

Published 2025-03-25 by Seth Larson
Reading time: 1 minute ♥ • 22

Whenever I talk about generative AI slop being sent into every conceivable communication platform I see a common suggestion on how to stop the slop from reaching human eyes:

"Just use AI to detect the AI"

We're already seeing companies offer this arrangement as a service. Just a few days ago Cloudflare announced they would use generative AI to create an infinite "labyrinth" for trapping AI

Sponsor me on GitHub
Follow me: @sethl
Keep up: stickmail
Email: weekly-reports

MAY 2025

CURL AND LIBCURL

DEATH BY A THOUSAND

RECENT POSTS

Hello Sponsor
July 28, 2025

EU-STF for funding critical Open Source
July 23, 2025

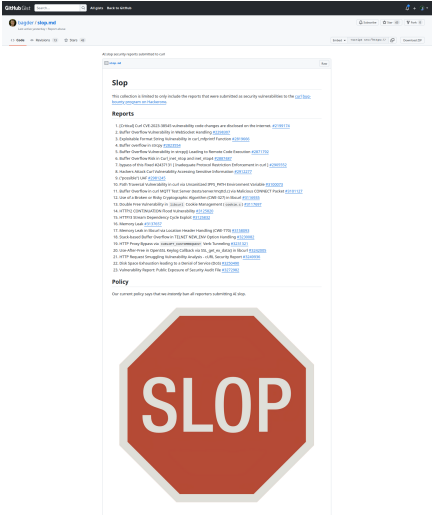
curl 8.15.0
July 16, 2025

Death by a thousand slops
July 14, 2025

I love I do it
July 13, 2025

Sponsor my laptop!
July 12, 2025

A Gist of Horrors



My data

- Hundreds of comments across more than 40 sites, posts, etc.

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev
- Social media and discussion platforms

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev
- Social media and discussion platforms

  LinkedIn, Slashdot, HackerNews

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev
- Social media and discussion platforms

  LinkedIn, Slashdot, HackerNews
- Repository and platform traces

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev
- Social media and discussion platforms

  LinkedIn, Slashdot, HackerNews
- Repository and platform traces

  HackerOne, GitHub

My data

- Hundreds of comments across more than 40 sites, posts, etc.
- Blog posts and podcasts

  Maintainers, Security Boulevard, OpenSourceSecurity.io
- News and trade articles

  The Register, The New Stack, Ars Technica, IT Pro, Pivot To AI, It's Foss, OpenTools.AI, Socket.Dev
- Social media and discussion platforms

  LinkedIn, Slashdot, HackerNews
- Repository and platform traces

  HackerOne, GitHub
- Document list initially developed from observation, supplemented by following links and doing searches

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?
- Tag data with semantic codes

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?
- Tag data with semantic codes
- “What’s the meaning of this?”

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?
- Tag data with semantic codes
- “What’s the meaning of this?”
- Grouping and re-grouping codes into themes

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?
- Tag data with semantic codes
- “What’s the meaning of this?”
- Grouping and re-grouping codes into themes
- “How are these ideas connected?”

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

Method: Thematic Analysis

Goal: Read textual data deeply to make sense of different lines of thought and argument.

- How does it work?
- Tag data with semantic codes
- “What’s the meaning of this?”
- Grouping and re-grouping codes into themes
- “How are these ideas connected?”
- Describing the themes you see

Braun, Virginia and Clarke, Victoria. (2008) Using thematic analysis in psychology. *Qualitative Research in Psychology*

From this...

About • Blog • RSS • Blogroll

Don't bring slop to a slop fight

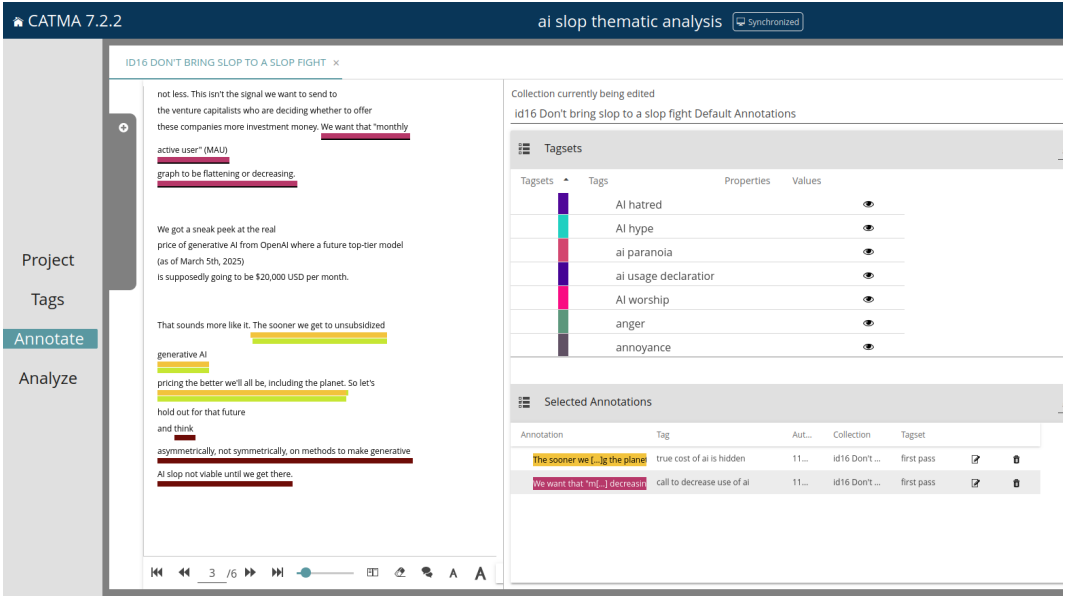
Published 2025-03-25 by Seth Larson
Reading time: 1 minute ♥ × 22

Whenever I talk about generative AI slop being sent into every conceivable communication platform I see a common suggestion on how to stop the slop from reaching human eyes:

> “Just use AI to detect the AI”

We're already seeing companies offer this arrangement as a service. Just a few days ago Cloudflare announced they would use generative AI to create an infinite "labyrinth" for trapping AI crawlers in pages of content and links.

To this:


CATMA 7.2.2
ai slop thematic analysis
Synchronized
Project
Tags
Annotate
Analyze
ID16 DON'T BRING SLOP TO A SLOP FIGHT
not less. This isn't the signal we want to send to
the venture capitalists who are deciding whether to offer
these companies more investment money. We want that "monthly
active user" (MAU)
graph to be flattening or decreasing.
We got a sneak peek at the real
price of generative AI from OpenAI where a future top-tier model
(as of March 5th, 2025)
is supposedly going to be $20,000 USD per month.
That sounds more like it. The sooner we get to unsubsidized
generative AI
pricing the better we'll all be, including the planet. So let's
hold out for that future
and think
asymmetrically, not symmetrically, on methods to make generative
AI slop not viable until we get there.
3 /6
Collection currently being edited
id16 Don't bring slop to a slop fight Default Annotations
Tagsets
Tagsets
Tags
Properties
Values
AI hatred
AI hype
ai paranoia
ai usage declaratior
AI worship
anger
annoyance
Selected Annotations
Annotation
Tag
Aut...
Collection
Tagset
The sooner we [...]g the plane
true cost of ai is hidden
11...
id16 Don't ...
first pass
We want that "m[...] decreasin
call to decrease use of ai
11...
id16 Don't ...
first pass

[What] So what even is this plausible slop?

- High volume, unwanted

[What] So what even is this plausible slop?

- High volume, unwanted
- Cheap to produce

[What] So what even is this plausible slop?

- High volume, unwanted
- Cheap to produce
- AI generated (context-sensitive, probabilistic)

[What] So what even is this plausible slop?

- High volume, unwanted
- Cheap to produce
- AI generated (context-sensitive, probabilistic)
- Difficult to validate and assess (but ultimated fabricated)

[What] So what even is this plausible slop?

- High volume, unwanted
- Cheap to produce
- AI generated (context-sensitive, probabilistic)
- Difficult to validate and assess (but ultimated fabricated)
- Attacks both social and technical structures

How bad is it?

- Good question.

How bad is it?

- Good question.
- Bad enough for maintainers to raise it

How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system

How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system
  - Lesson from XZ vulnerability: Attacks targeting maintainer capacity can cause damage

How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system
  - Lesson from XZ vulnerability: Attacks targeting maintainer capacity can cause damage
  - Automation allows for fast iteration / mutation

How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system
  - Lesson from XZ vulnerability: Attacks targeting maintainer capacity can cause damage
  - Automation allows for fast iteration / mutation
- Already quite sophisticated and it's early days.

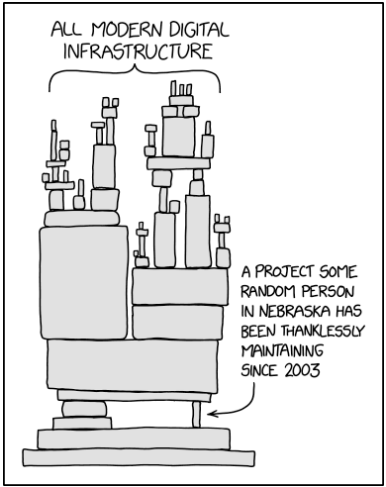
How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system
  - Lesson from XZ vulnerability: Attacks targeting maintainer capacity can cause damage
  - Automation allows for fast iteration / mutation
- Already quite sophisticated and it's early days.
- If detected: Draws time away from real security issues

How bad is it?

- Good question.
- Bad enough for maintainers to raise it
- Strain to an already strained system
  - Lesson from XZ vulnerability: Attacks targeting maintainer capacity can cause damage
  - Automation allows for fast iteration / mutation
- Already quite sophisticated and it's early days.
- If detected: Draws time away from real security issues
- If not detected: Waste, fraud, abuse, malware

How bad does it have to be?



[Dependency. Randall Monroe. https://xkcd.com/2347, CC-SA-NC 2.5]

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them
  - Using AI where once we had genuine interactions ("dead Internet theory")

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them
  - Using AI where once we had genuine interactions ("dead Internet theory")
  - Breakdowns in pathways of newcomers joining

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them
  - Using AI where once we had genuine interactions ("dead Internet theory")
  - Breakdowns in pathways of newcomers joining
  - Burnout for contributors and leaders

How bad is it?

What's our threat model here? How would we go about measuring the problem?

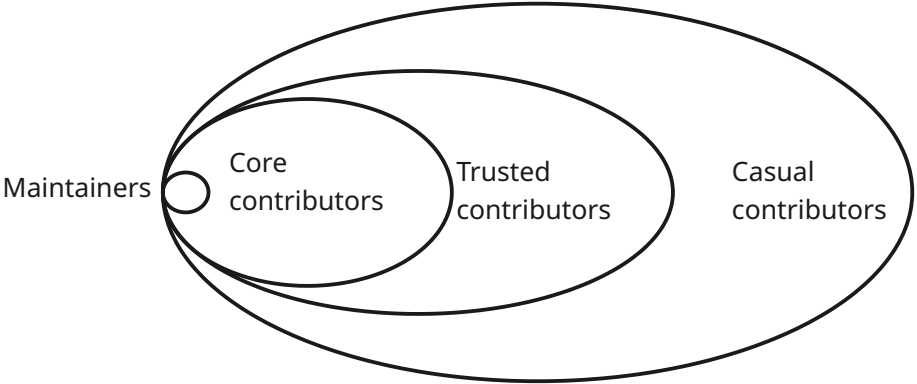
- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them
  - Using AI where once we had genuine interactions ("dead Internet theory")
  - Breakdowns in pathways of newcomers joining
  - Burnout for contributors and leaders
  - Worst case scenario: "Social model collapse"

How bad is it?

What's our threat model here? How would we go about measuring the problem?

- Slop drowning out real problems
- Waste, fraud, abuse
- Social structures under strain
  - Treating people like AI discourages them
  - Using AI where once we had genuine interactions ("dead Internet theory")
  - Breakdowns in pathways of newcomers joining
  - Burnout for contributors and leaders
  - Worst case scenario: "Social model collapse"
    ...what's this social model?

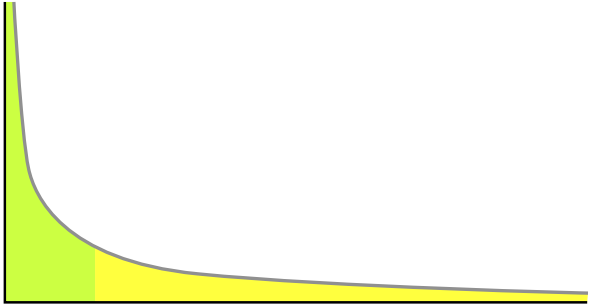
Onion model



[Not to scale]

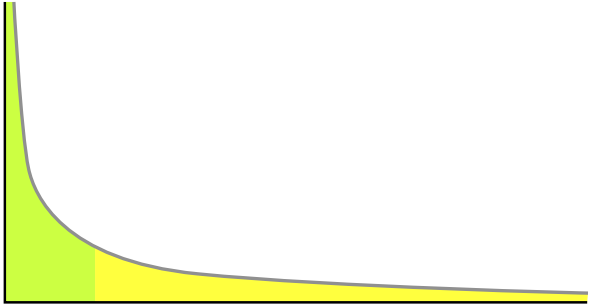
Long tail model

- 80-20 rule



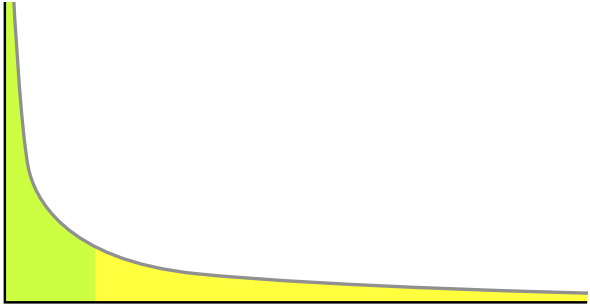
Long tail model

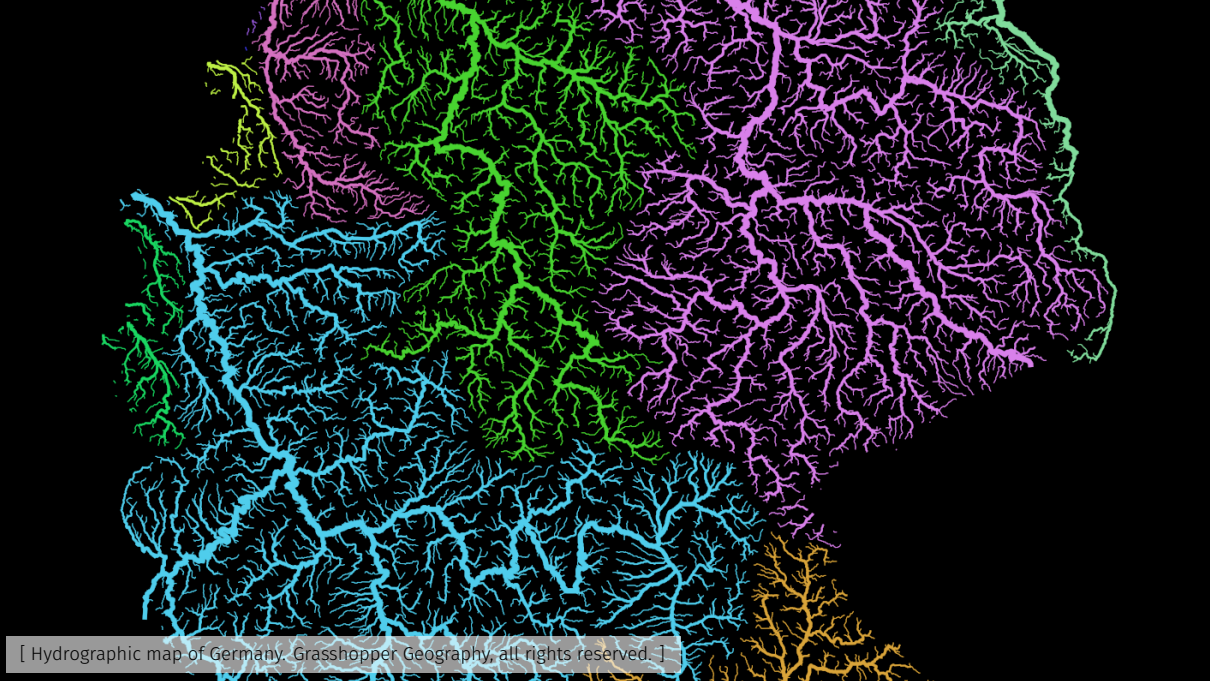
- 80-20 rule
- ...might feel more like the 99.5-.5 rule...



Long tail model

- 80-20 rule
- ...might feel more like the 99.5-.5 rule...
- Core work from a smaller group, unique work from an expanded group





[What] So what even is this plausible slop?

Recollect definition:

- High volume, unwanted
- Cheap to produce
- AI generated (context-sensitive, probabilistic)
- Difficult to validate and assess (but ultimately fabricated)
- Attacks both social and technical structures

Is slop like spam?

Spam:

- High volume, unwanted messages ✓

Is slop like spam?

Spam:

- High volume, unwanted messages ✓
- Cheap to produce ✓

Is slop like spam?

Spam:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗

Is slop like spam?

Spam:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗
- Easy to detect ✗

Is slop like spam?

Spam:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗
- Easy to detect ✗

- Attacks (some social?) certainly technical structures ⇔

Is slop like a denial of service attack?

A DDOS is:

- High volume, unwanted messages ✓

Is slop like a denial of service attack?

A DDOS is:

- High volume, unwanted messages ✓
- Cheap to produce ✓

Is slop like a denial of service attack?

A DDOS is:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗

Is slop like a denial of service attack?

A DDOS is:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗
- Easy to detect ✗

Is slop like a denial of service attack?

A DDOS is:

- High volume, unwanted messages ✓
- Cheap to produce ✓
- Procedurally generated ✗
- Easy to detect ✗
- Attacks technical structures ⇔

Is slop like spearphishing?

Spearphishing is:

- Low volume messages ✗

Is slop like spearphishing?

Spearphishing is:

- Low volume messages ✗
- Challenging to produce ✗

Is slop like spearphishing?

Spearphishing is:

- Low volume messages ✗
- Challenging to produce ✗
- Custom generated ✓

Is slop like spearphishing?

Spearphishing is:

- Low volume messages ✗
- Challenging to produce ✗
- Custom generated ✓
- Hard to detect ✓

Is slop like spearphishing?

Spearphishing is:

- Low volume messages ✗
- Challenging to produce ✗
- Custom generated ✓
- Hard to detect ✓
- Attacks both social and technical structures ✓

Slop vs historical examples

	Volume	Production	Toolset	Identification	Structural Cost
Spam	High	Easy	Procedural	Easy	(Social?) & Technical
DDOS	High	Easy	Procedural	Easy	Technical
Spearphishing	Low	Hard	Augmented	Hard	Social & Technical

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation
  - May improve in their skill at wielding AI

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation
  - May improve in their skill at wielding AI
- Are these attackers?

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation
  - May improve in their skill at wielding AI
- Are these attackers?
  - May just be playing the odds

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation
  - May improve in their skill at wielding AI
- Are these attackers?
  - May just be playing the odds
  - May mutate to improve their strategy

Why are people doing this?

- Are these well-intended newcomers whom we might find a way to welcome?
  - Is slop like an eternal September?
  - Able to be socialized into the community
  - Discouraged if treated with hostility
  - An investment in sustainability
  - ....if such newcomers are indeed who is authoring the slop.
- Are these reward-seekers? (bounties, glory)
  - Likely to respond to enforcement and consistency: rule changes, validation routines
  - Perhaps less concerned with social incorporation
  - May improve in their skill at wielding AI
- Are these attackers?
  - May just be playing the odds
  - May mutate to improve their strategy
- And if we can't adjust to the volume, does it matter why?

Dilemma

- Current approaches are insufficient

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules

    ...except for the rules about telling the truth...

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules

    ...except for the rules about telling the truth...
- My take: Ad hoc ‘know it when I see it’ detection also seems likely to be insufficient

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules

    ...except for the rules about telling the truth...
- My take: Ad hoc ‘know it when I see it’ detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc ‘know it when I see it’ detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc ‘know it when I see it’ detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.
  - Thanks to generative AI tools, scammers can now have excellent grammar and correct contextual information!

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc 'know it when I see it' detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.
  - Thanks to generative AI tools, scammers can now have excellent grammar and correct contextual information!
  - Probably not reasonable to interpret language fluidity as safety

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc 'know it when I see it' detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.
  - Thanks to generative AI tools, scammers can now have excellent grammar and correct contextual information!
  - Probably not reasonable to interpret language fluidity as safety
- As newcomers seek to be better, they may follow rules more closely and look more like AI over time.

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc 'know it when I see it' detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.
  - Thanks to generative AI tools, scammers can now have excellent grammar and correct contextual information!
  - Probably not reasonable to interpret language fluidity as safety
- As newcomers seek to be better, they may follow rules more closely and look more like AI over time.
- Model 'tells' are being ironed away rapidly.

Dilemma

- Current approaches are insufficient
- My take: Governance is also likely to be insufficient
  - Slop is well-formatted and follows the rules
    ...except for the rules about telling the truth...
- My take: Ad hoc 'know it when I see it' detection also seems likely to be insufficient
  - Recall the now-dated advice about phishing and scams...
    ...to look for bad grammar and mismatched information.
  - Thanks to generative AI tools, scammers can now have excellent grammar and correct contextual information!
  - Probably not reasonable to interpret language fluidity as safety
- As newcomers seek to be better, they may follow rules more closely and look more like AI over time.
- Model 'tells' are being ironed away rapidly.
- Fraudsters and attackers also get better and smarter – they might look less like AI over time with better prompt engineering

How can we respond?

- I think we need to draw from both our trusted arsenal and build up new strategies

How can we respond?

- I think we need to draw from both our trusted arsenal and build up new strategies
- This is an active area of work!

How can we respond?

- I think we need to draw from both our trusted arsenal and build up new strategies
- This is an active area of work!
- You've seen some of my analysis of the problem

How can we respond?

- I think we need to draw from both our trusted arsenal and build up new strategies
- This is an active area of work!
- You've seen some of my analysis of the problem
- I'm drawing heavily from the community

How are communities responding to this threat?

- In action right now:

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers
  - Rules & governance

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers
  - Rules & governance
    - AI disclosure requirements

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers
  - Rules & governance
    - AI disclosure requirements
    - AI bans

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers
  - Rules & governance
    - AI disclosure requirements
    - AI bans
  - Education & awareness

How are communities responding to this threat?

- In action right now:
  - Support & encouragement for maintainers
  - Rules & governance
    - AI disclosure requirements
    - AI bans
  - Education & awareness
- Dialectical oppositions in proposed solutions

don't bring slop to a slop fight	⟷	AI to stop AI
blame tool not user	⟷	blame user not tool
need for automation	⟷	need for education
platform withdrawal	⟷	platform improvements
collective strategies	⟷	per-project strategies
alternate paradigms financial bonds trust webs	⟷	improvements to current strategies faster detection raise barriers to submission seek regulation

Next steps:

- Continuing my cataloging, but shifting gears to quantitative work

Next steps:

- Continuing my cataloging, but shifting gears to quantitative work
- Measuring impacts (both of slop and counter measures)

Next steps:

- Continuing my cataloging, but shifting gears to quantitative work
- Measuring impacts (both of slop and counter measures)
- Prototyping responses

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:
  - Must use a toolchain consistent with software freedom:

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:
  - Must use a toolchain consistent with software freedom: open source, open data, open models, open weights.

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:
  - Must use a toolchain consistent with software freedom: open source, open data, open models, open weights.
  - Use “as small as possible” language models (lightweight to train and run)

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:
  - Must use a toolchain consistent with software freedom:
    open source, open data, open models, open weights.
  - Use “as small as possible” language models (lightweight to train and run)
  - Collectively available but locally tuned

Contours of a response

- Governance responses collection: who is making what policy moves, and is it helping?
- Building models to support automated approaches:
  - Must use a toolchain consistent with software freedom: open source, open data, open models, open weights.
  - Use “as small as possible” language models (lightweight to train and run)
  - Collectively available but locally tuned
- Beyond detection: Improved validation tools

What do you think?

Are you seeing slop in your part of FLOSS? How plausible is it?

What kinds of traits would you look for in a solution?

Questions? Feedback?

kaylea@uw.edu—@kaylea@social.coop

https://kayleachampion.com

https://communitydata.science—@communitydata@social.coop

This work has been unfunded so far. I am actively seeking new funding and collaborators!

[Washington Coast. John Murphy. (kingair42) via flickr, CC BY-SA 2.0]

Why are people doing this in the first place?

- Newcomers

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory
  - high quality submissions

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory
  - high quality submissions
- Fraudsters, attackers

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory
  - high quality submissions
- Fraudsters, attackers
  - payment

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory
  - high quality submissions
- Fraudsters, attackers
  - payment
  - glory

Different motives suggest different responses.

Why are people doing this in the first place?

- Newcomers
  - good faith contributors
  - range of motives
  - low quality submissions
- Experienced reward seekers
  - good faith contributors
  - payment
  - glory
  - high quality submissions
- Fraudsters, attackers
  - payment
  - glory
  - deceptive submissions

Different motives suggest different responses.