



# Human AI Interaction

Lecture 18: Saying thanks!  
[aidesignclass.org](http://aidesignclass.org)

# Learning goals today

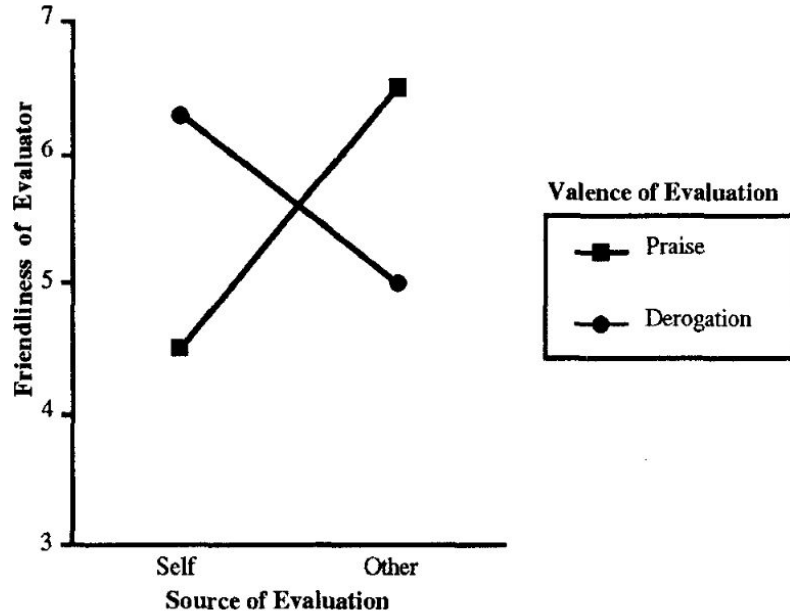
- Distinguish situations where computers are seen as social actors, and where they are not
  - People often think of computers as social actors, and trust them similarly
  - But robots/computers are not seen as human always
- You can't really be grateful to a computer
  - Why not?
- Credit, attribution, and data
  - While computers can help with data provenance and attribution, credit is a different matter

# Robot under attack

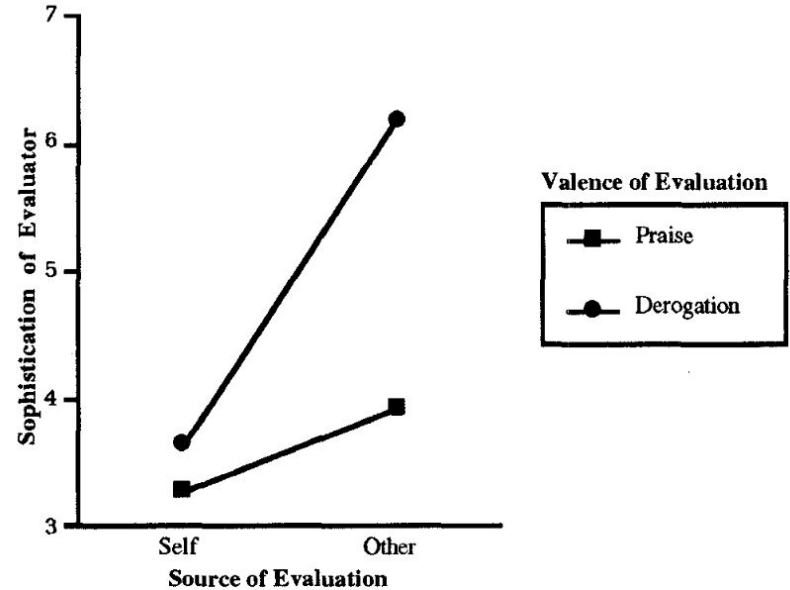
Q: how does this video make you feel?



# Computers as social actors



**Figure 2:** Perceived Friendliness as a Function of the Source and Valence of the Evaluation ( $N = 44$ )



**Figure 3:** Perceived Sophistication as a Function of the Source and Valence of the Evaluation ( $N = 44$ )

# Should you trust a robot?

Q: What are the implications of this study for designing a system like ChatGPT?



# Can computers say “thank you”?

Most of what we speak about today is inspired by this paper

Monroy-Hernández, Andrés, et al. "[Computers can't give credit: How automatic attribution falls short in an online remixing community.](#)" *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011.



**Figure 2. Screenshot of a Scratch remix project highlighting automatic (the area inside the bottom left circle) and manual (the area in the top right circle) attribution.**

# Can computers say “thank you”?

“Hi i’m koolkid15 the original creator of luigi disco jay-

man41 copied me!! and didn’t even aknowladge me he

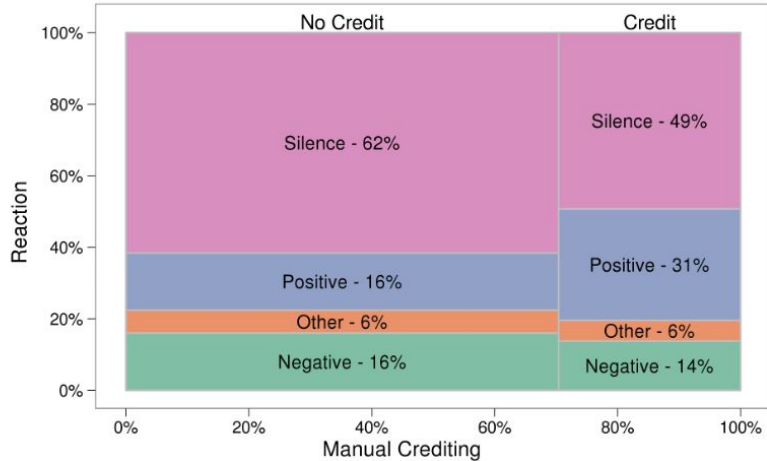
didn’t change anything !! I wrote or drew!! and jay-

man...if your reading this think about other people!!!!”



**Figure 2. Screenshot of a Scratch remix project highlighting automatic (the area inside the bottom left circle) and manual (the area in the top right circle) attribution.**

# Can computers say “thank you”?



**Figure 6.** Mosaic plot showing the distribution of reactions of original creators who had viewed remixes of their project and indicating whether they left manual credit. The proportion of response types is shown along the y-axis. The proportion of projects including and omitting manual credit is shown along the x-axis. ( $n = 932$ )



**Figure 2.** Screenshot of a Scratch remix project highlighting automatic (the area inside the bottom left circle) and manual (the area in the top right circle) attribution.



# Removing your data

See [this form](#)

- What are your reactions about this based on what you know about credit?

### OpenAI Personal Data Removal Request

Under certain privacy or data protection laws, such as the GDPR, you may have the right to object to the processing of your personal data by OpenAI's models. You can submit that request using this form.

Please provide complete, accurate, and relevant answers on this form for evaluation. OpenAI may use additional sources to verify information, balancing privacy and free expression in accordance with applicable law. Submitting a request does not guarantee that information about you will be removed from ChatGPT outputs, and incomplete forms may not be processed.

Read this [Help Center article](#) for more about how we collect and use personal data to develop ChatGPT.

#### Your information

Please provide your own full legal name, even if you are making the request on behalf of someone else. If you are submitting the request on behalf of someone else, you must have the legal authority to act on their behalf.

First name \*  Last name \*

Email \*   
Please provide an email address we can use to contact you about your request.

Country whose law applies \*   
Please select the country whose law applies. Typically your country of residence.

I am acting on behalf of... \*

#### Legal Relationship

Please describe your legal relationship to the person on whose behalf this request is being made.

#### Data subject's information

Please provide the full legal name for the subject of the request (the data subject). If you are completing this request on behalf of yourself, please provide your own full legal name.

Data subject's first name \*  Data subject's last name \*

Public figure \*

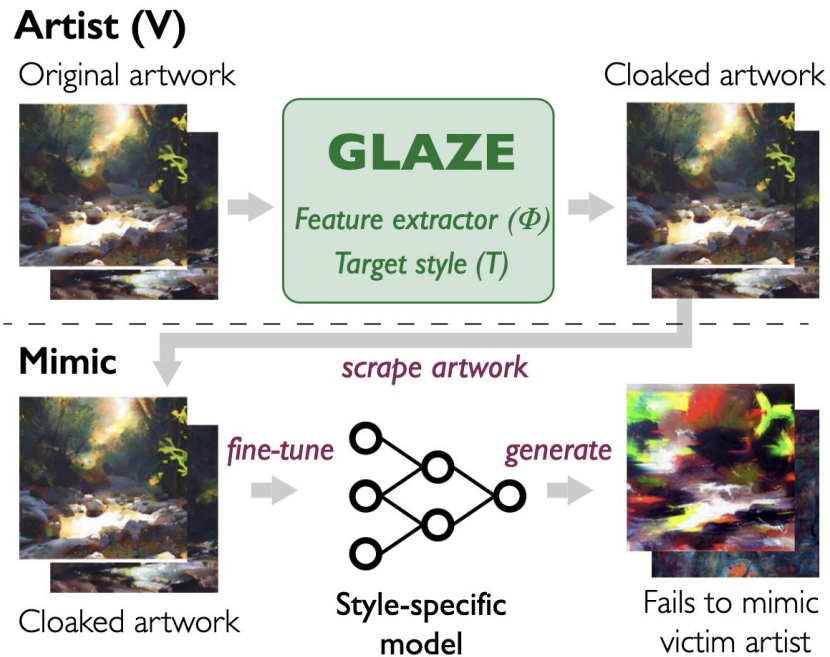
# AI Resistance

“Many artists (> 89% artists) have already or plan to take actions because of AI mimicry. Over 95% of artists post their artwork online. Out of these artists, 53% of them anticipate reducing or removing their online artwork, if they haven’t already. Out of these artists, 55% of them believe reducing their online presence will significantly impact their careers. One participant stated “AI art has unmotivated myself from uploading more art and made me think about all the years I spent learning art.” 78% of artists anticipate AI mimicry would impact their job security, and this percentage increases to 94% for the job security of newer artists. Further, 24% of artists believe AI art has already impacted their job security, and an additional 53% expect to be affected within the next 3 years.”

From <https://arxiv.org/pdf/2302.04222.pdf>

# Glaze

Key insight: if you know what features are being transferred/copied, you can perturb just those features



**Figure 5.** Overview of *Glaze*, a system that protects victim artists from AI style mimicry by cloaking their online artwork. **(Top)** An artist  $V$  applies the cloaking algorithm (uses a feature extractor  $\Phi$  and a target style  $T$ ) to generate cloaked versions of  $V$ 's art pieces. Each cloak is a small perturbation unnoticeable to human eye. **(Bottom)** A mimic scrapes the cloaked art pieces from online and uses them to fine-tune a model to mimic  $V$ 's style. When prompted to generate artwork in the style of  $V$ , mimic's model will generate artwork in the target style  $T$ , rather than  $V$ 's true style.

# Credit vs. Resistance vs. Removal

- When would you use each method?
- The answers are socio-technical, not just technical

POLL:

REMOTE class next tuesday?