

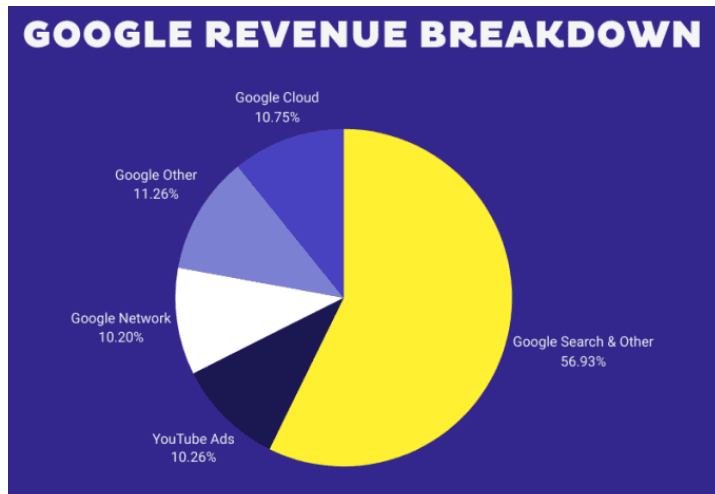
CMSC 3540 I: The Interplay of Economics and ML
(Winter 2024)

The Economics of Online Content Creation
and Generative AIs

Instructor: Haifeng Xu

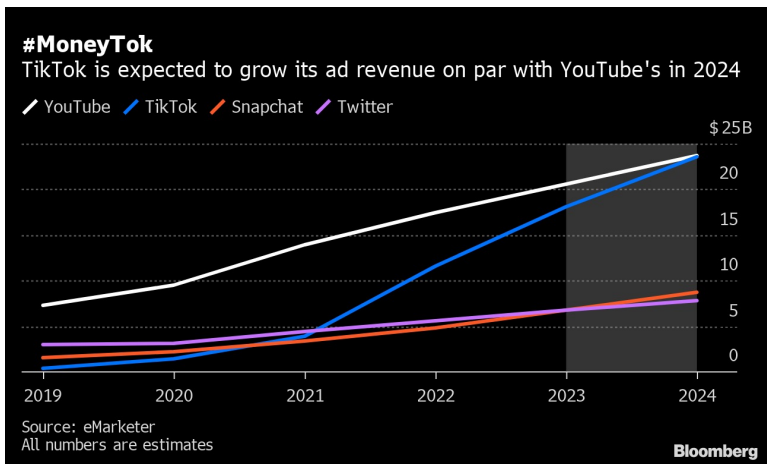
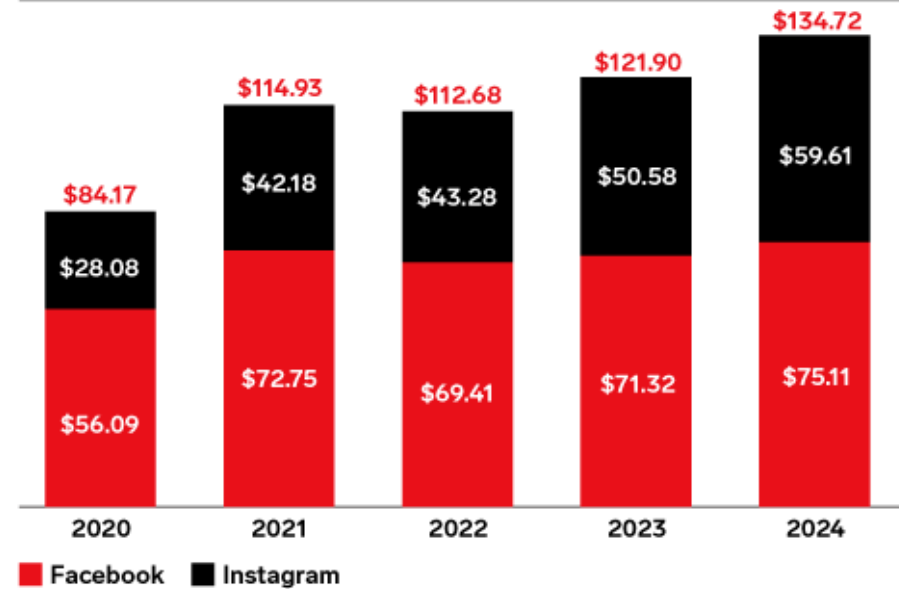


Online Content Creation is a Huge Economy



Meta Net Ad Revenues Worldwide, by Segment, 2020-2024

billions



Competition between Human-Generated and AI-Generated Contents is Already Here



backpropagation with rigorous formulation



CMU School of Computer Science

<http://www.cs.cmu.edu/~Werbos.backprop.pdf> PDF

Backpropagation through time

by PJ WERBOS · Cited by 6665 — Under certain conditions, this can be a **rigorous** approach but its generality is limited, and it requires great care with the side conditions (which are rare) 11 pages



Deep Learning @ VU

<https://dlvu.github.io/slides/dlvu.lecture02.pdf> PDF

Lecture 2: Backpropagation - Peter Bloem

Today's lecture will be entirely devoted to the **backpropagation** algorithm. The heart of all deep learning. part 1: review part 2: scalar **backpropagation** part 3 ... 46 pages



ScienceDirect.com

<https://www.sciencedirect.com/topics/computer-science>

Backpropagation Algorithm - an overview

The **backpropagation** algorithm performs learning on a multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class ...

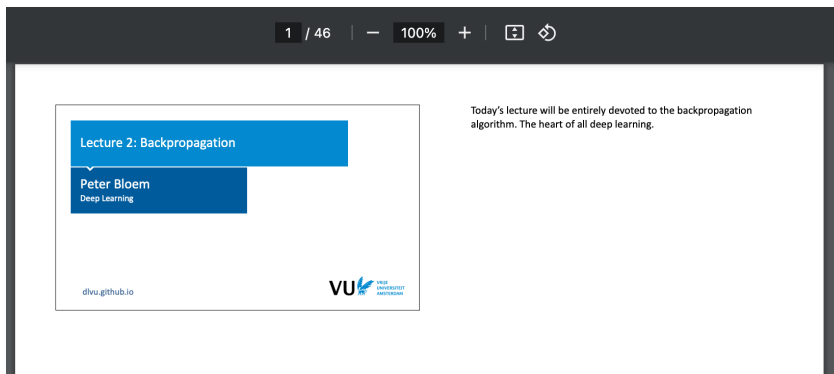


Medium · zfeng

1 like · 3 years ago

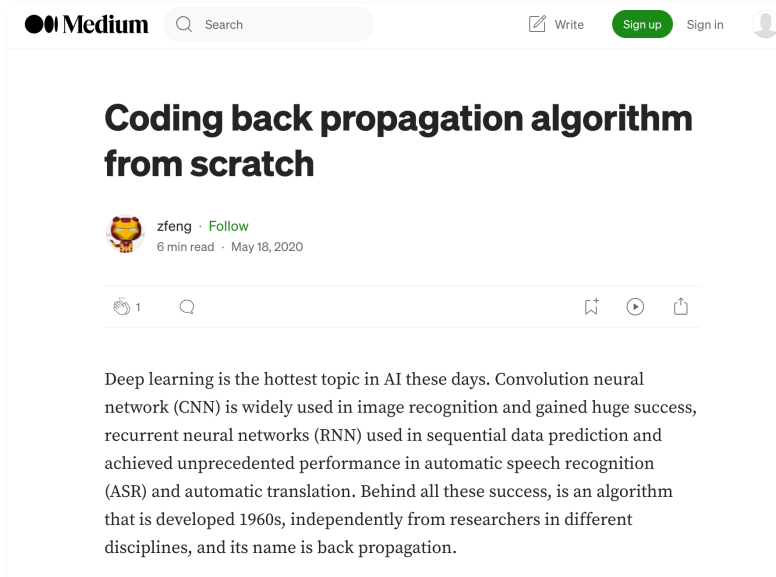
Coding back propagation algorithm from scratch | by zfeng

In this section I will define briefly how we can derive the **back propagation** algorithm. Here I only list the key steps in the derivation ...



Human created contents

Competition between Human-Generated and AI-Generated Contents is Already Here



The screenshot shows a Medium article page. At the top, the Medium logo and search bar are visible. The article title is "Coding back propagation algorithm from scratch" by user "zfeng". The article text begins with "Deep learning is the hottest topic in AI these days. Convolution neural network (CNN) is widely used in image recognition and gained huge success, recurrent neural networks (RNN) used in sequential data prediction and achieved unprecedented performance in automatic speech recognition (ASR) and automatic translation. Behind all these success, is an algorithm that is developed 1960s, independently from researchers in different disciplines, and its name is back propagation."

Another content created by human



backpropagation with rigorous formulation

CMU School of Computer Science
<http://www.cs.cmu.edu> › [Werbos.backprop.pdf](#) PDF

Backpropagation through time

by PJ WERBOS · Cited by 6665 — Under certain conditions, this can be a **rigorous** approach but its generality is limited, and it requires great care with the side conditions (which are rare) 11 pages

Deep Learning @ VU
<https://dlvu.github.io> › slides › [dlvu.lecture02.pdf](#) PDF

Lecture 2: Backpropagation - Peter Bloem

Today's lecture will be entirely devoted to the **backpropagation** algorithm. The heart of all deep learning. part 1: review part 2: scalar **backpropagation** part 3 ... 46 pages

ScienceDirect.com
<https://www.sciencedirect.com> › topics › computer-science

Backpropagation Algorithm - an overview

The **backpropagation** algorithm performs learning on a multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class ...

Medium · zfeng
1 like · 3 years ago

Coding back propagation algorithm from scratch | by zfeng




In this section I will define briefly how we can derive the **back propagation** algorithm. Here I only list the key steps in the derivation ...

Competition between Human-Generated and AI-Generated Contents is Already Here



Google

backpropagation with rigorous formulation

 CMU School of Computer Science
<http://www.cs.cmu.edu> › Werbos.backprop.pdf  

Backpropagation through time

by PJ WERBOS · Cited by 6665 — Under certain conditions, this can be a **rigorous** approach but its generality is limited, and it requires great care with the side conditions (which are rare) 11 pages

 Deep Learning @ VU
<https://dlvu.github.io> › slides › dlvu.lecture02.pdf  



Lecture 2: Backpropagation - Peter Bloem

Today's lecture will be entirely devoted to the **backpropagation** algorithm. The heart of all deep learning. part 1: review part 2: scalar **backpropagation** part 3 ... 46 pages

 ScienceDirect.com
<https://www.sciencedirect.com> › topics › computer-science 

Backpropagation Algorithm - an overview

The **backpropagation** algorithm performs learning on a multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class ...

 Medium · zfeng
1 like · 3 years ago 

Coding back propagation algorithm from scratch | by zfeng

In this section I will define briefly how we can derive the **back propagation** algorithm. Here I only list the key steps in the derivation ...

Competition between Human-Generated and AI-Generated Contents is Already Here



ChatGPT 4 ▾

HA You

Could you use explain backpropagation to me with rigorous formulations?

ChatGPT

Certainly! To explain backpropagation with a more rigorous formulation, we'll delve into the mathematics that underpin the algorithm. This involves understanding how the gradient of the loss function is computed with respect to the weights of the network and how these gradients are used to update the weights in a way that minimizes the loss.

Notation and Setup

Let's consider a simple feedforward neural network with L layers, where each layer l has $n^{[l]}$ neurons. The network takes an input x and produces an output \hat{y} , aiming to approximate the true output y . Each layer l has a weight matrix $W^{[l]}$, a bias vector $b^{[l]}$, and an activation function $g^{[l]}$.

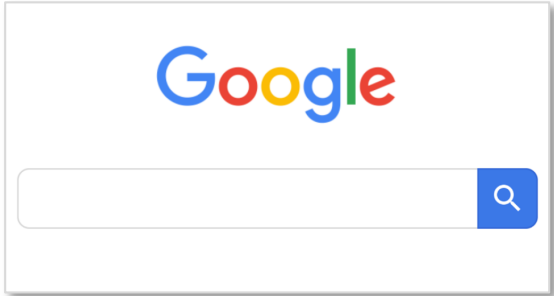
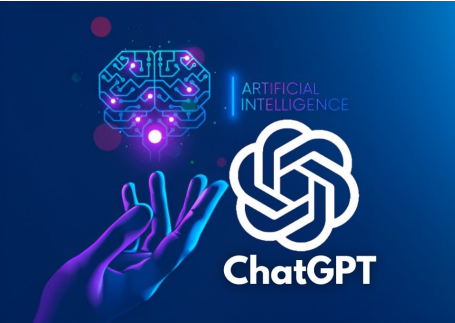
Forward Pass

1. **Input Layer:** The input layer receives the input $x = a^{[0]}$.
2. **Hidden Layers and Output Layer:** For each layer l from 1 to L , the pre-activation value $z^{[l]}$ and the activation $a^{[l]}$ are computed as follows:

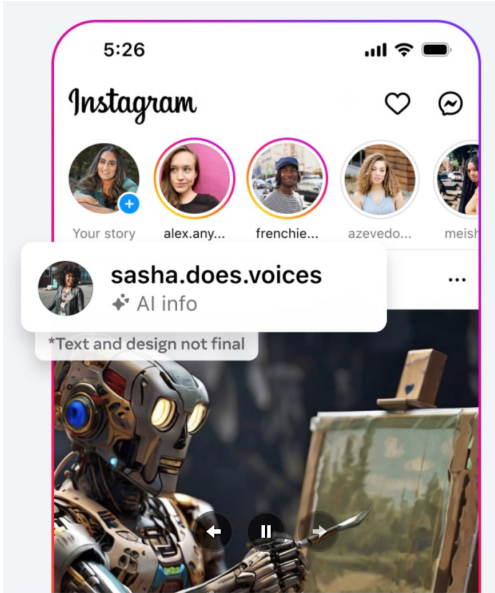
$$z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]}$$

$n^{[l-1]} \times n^{[l]}$

Competition between Human-Generated and AI-Generated Contents is Already Here

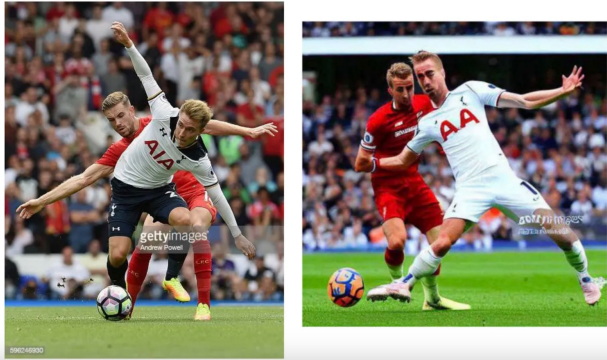


Competition between Human-Generated and AI-Generated Contents is Already Here



Competition between Human-Generated and AI-Generated Contents is Already Here

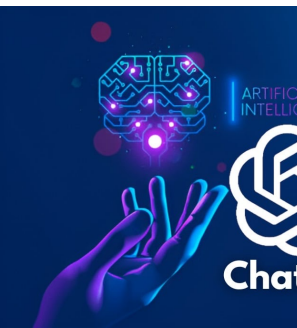
Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement



The competition not only affects the Internet ecosystem, but also has real-world consequences

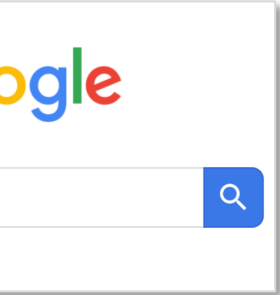


Competition between Human-Generated and AI-Generated Contents is Already Here



Creating video from text

Sora is an AI model that can create realistic and imaginative scenes from text instructions.

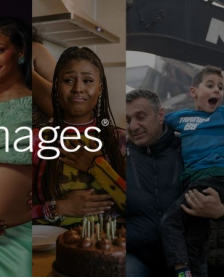


Texts



Prompt: Several giant wooly mammoths approach treading through a snowy meadow, their long wooly fur lightly blows in the wind as they walk, snow covered trees and dramatic sno... more

0:00 / 0:10

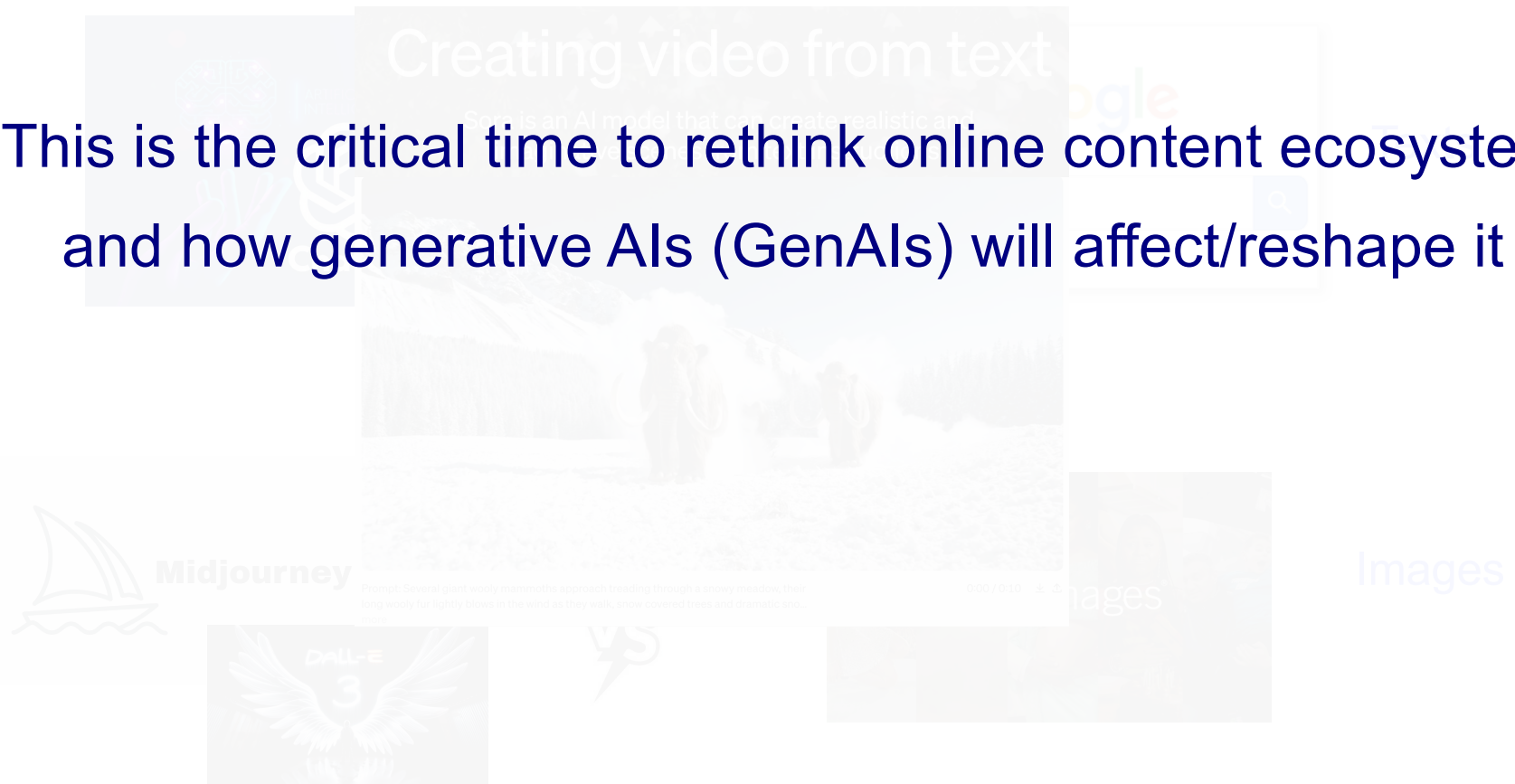


Images



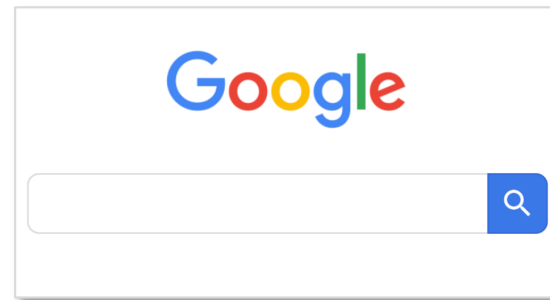
Competition between Human-Generated and AI-Generated Contents is Already Here

This is the critical time to rethink online content ecosystem and how generative AIs (GenAIs) will affect/reshape it



Two Modes that GenAIs Can Affect Content Creation

1. GenAIs **as representations** of an Internet content ecosystem



Texts

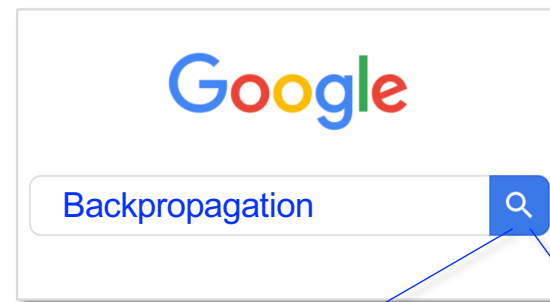
2. GenAIs **create part** of an Internet content ecosystem



Images

Two Modes that GenAIs Can Affect Content Creation

1. GenAIs as representations of an Internet content ecosystem



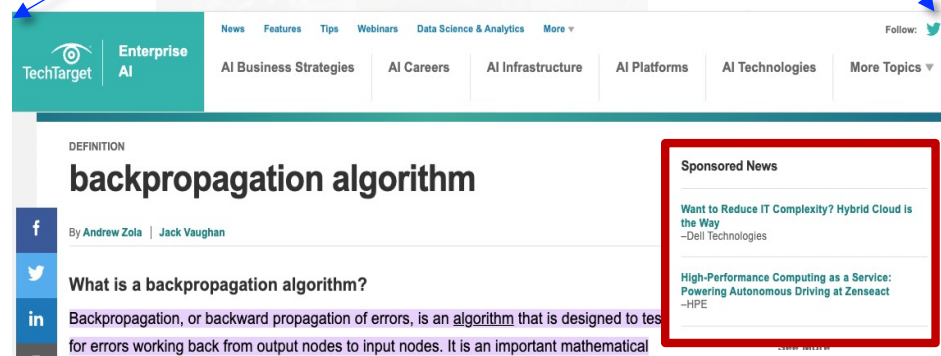
Texts

Sink of user traffic

Bridges user traffic and online contents

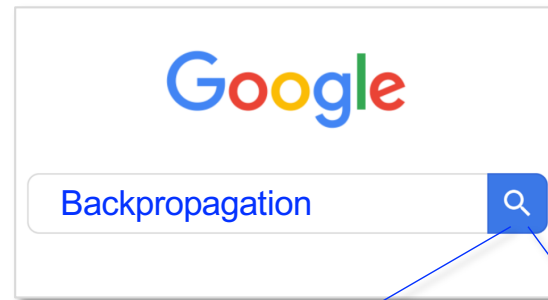
Directly resolves queries

Users have no need to visit content websites any more



Two Modes that GenAIs Can Affect Content Creation

1. GenAIs as representations of an Internet content ecosystem



Texts

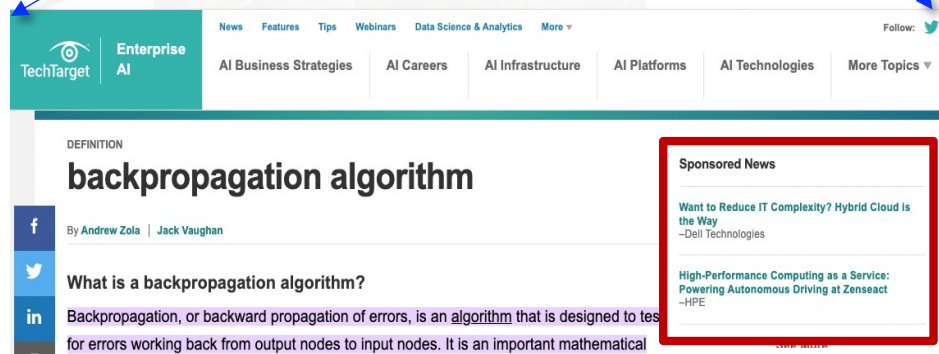
How would such GenAI technology affect content creation?

Sink of user traffic

Bridges user traffic and online contents

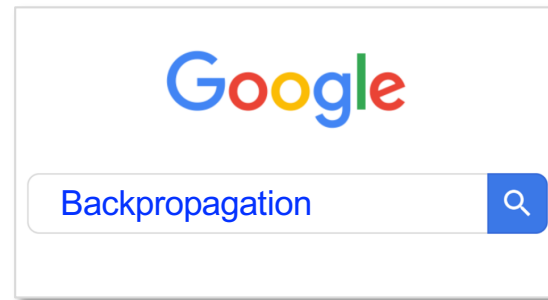
Directly resolves queries

Users have no need to visit content websites any more



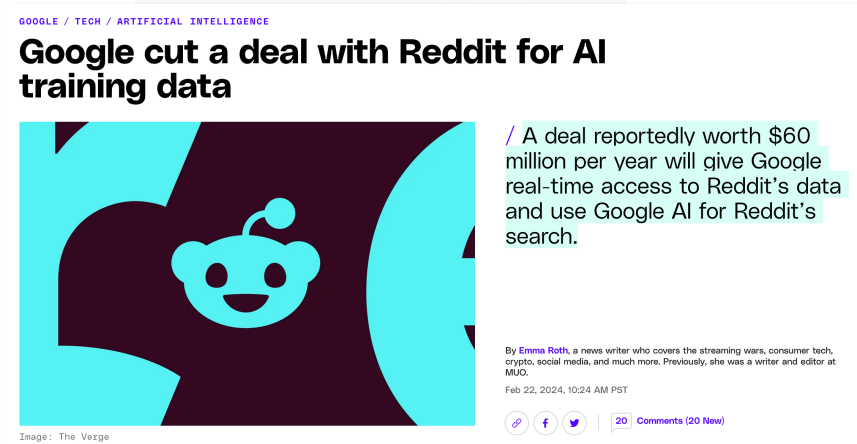
Two Modes that GenAIs Can Affect Content Creation

1. GenAIs as representations of an Internet content ecosystem



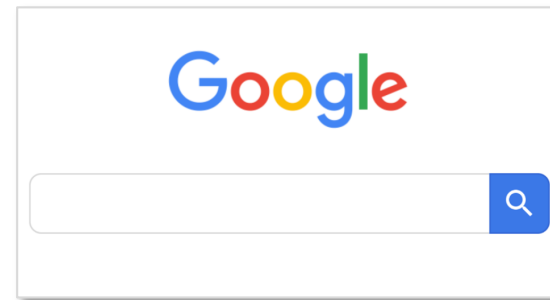
Texts

How would such GenAI technology affect content creation?



Two Modes that GenAIs Can Affect Content Creation

1. GenAIs **as representations** of an Internet content ecosystem



Texts

2. GenAIs **create part** of an Internet content ecosystem



Images

Reminder of This Lecture

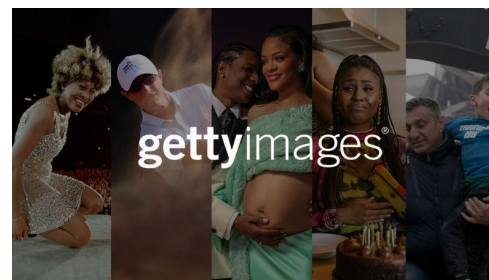
...will zoom into the second mode

1. GenAIs as representations of an Internet content ecosystem



Texts

2. GenAIs **create part** of an Internet content ecosystem



Images

Reminder of This Lecture

...will zoom into the second mode

1. GenAIs as representations of an Internet content ecosystem

- Tries to answer one economic question – *how AI content generation affects incentives of human creators in recommender systems and the equilibrium?*
- [Link to a relevant recent workshop at AAI 2024](#)



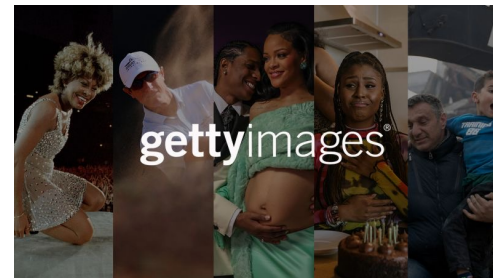
Credit: DALL.E

Texts

2. GenAIs **create part** of an Internet content ecosystem



VS



Images

Human vs. Generative AI in Content Creation Competition: Symbiosis or Conflict?

Fan Yao¹

FY4BC@VIRGINIA.EDU

Chuanhao Li²

CHUANHAO.LI.CL2637@YALE.EDU

Denis Nekipelov^{1,3}

DN4W@VIRGINIA.EDU

Hongning Wang⁴

WANG.HONGN@GMAIL.COM

Haifeng Xu⁵

HAIFENGXU@UCHICAGO.EDU

¹*Department of Computer Science, University of Virginia, USA*

²*Department of Statistics and Data Science, Yale University, USA*

³*Department of Economics, University of Virginia, USA*

⁴*Department of Computer Science and Technology, Tsinghua University, China*

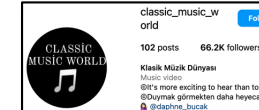
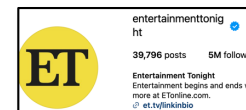
⁵*Department of Computer Science, University of Chicago, USA*

Paper link: <https://arxiv.org/pdf/2402.15467.pdf>

Built upon recent *multi-agent system modeling* for content creation competition among human creators [BT'17, HKJKD'22, YLNWX'22, JGS'23]

Utility-maximizing human content creators

Creator i



Attracts $\frac{x_{ik}}{\sum_j x_{jk}}$ of topic k users

create $x_{ik} \in \mathbb{R}_+$ amount of contents for topic k



Topic 1



μ_1



Topic 2



μ_2



Topic 3



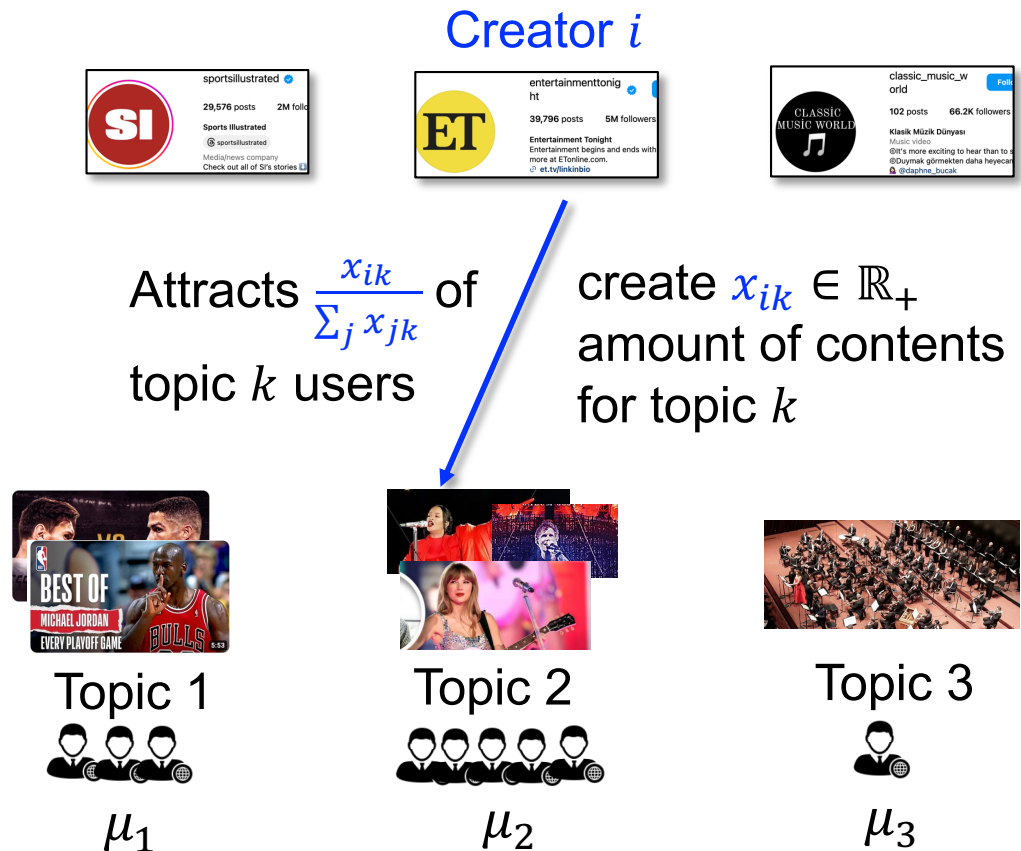
μ_3

Size of user population

Built upon recent *multi-agent system modeling* for content creation competition among human creators [BT'17, HKJKD'22, YLNWX'22, JGS'23]

Payoff of creator $i =$

$$\sum_k \frac{x_{ik} \cdot \mu_k}{\sum_j x_{jk}} - c_i(\mathbf{x}_i)$$

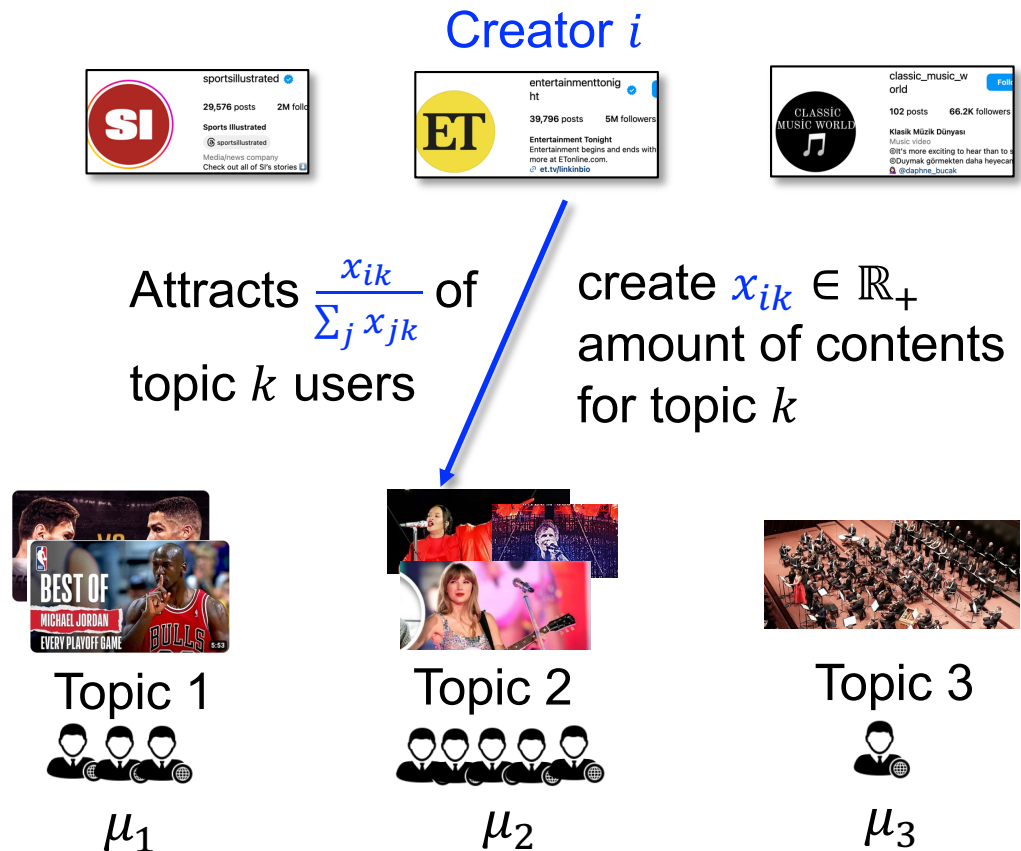


Built upon recent *multi-agent system modeling* for content creation competition among human creators [BT'17, HKJKD'22, YLNWX'22, JGS'23]

Payoff of creator $i =$

$$\sum_k \frac{x_{ik} \cdot \mu_k}{\sum_j x_{jk}} - c_i(\mathbf{x}_i)$$

- i 's user traffic from content topic k
- Captures competition (i.e., Tullock contest)

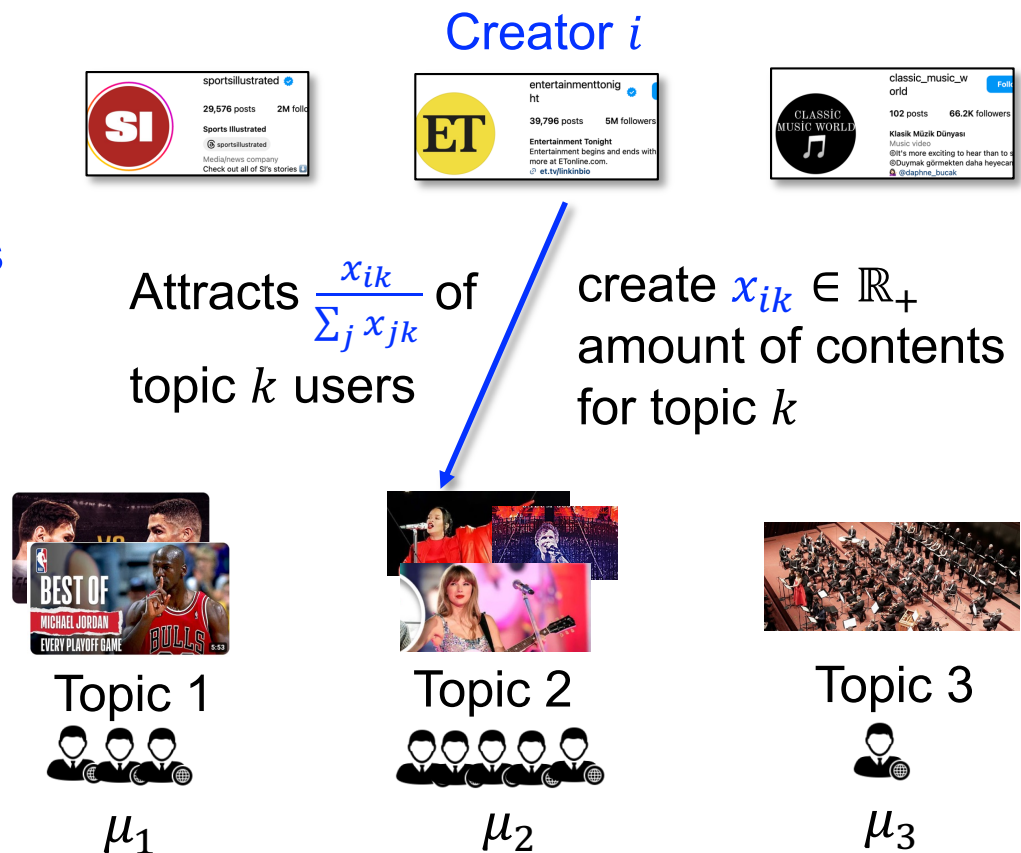


Built upon recent *multi-agent system modeling* for content creation competition among human creators [BT'17, HKJKD'22, YLNWX'22, JGS'23]

Payoff of creator $i =$

$$\sum_k \frac{x_{ik} \cdot \mu_k}{\sum_j x_{jk}} - c_i(\mathbf{x}_i)$$

- i 's cost of creating contents of different topics
- c_i captures i 's expertise (i.e., $c_i(\mathbf{x}) = x_1^2 + 0.1x_2^2$)

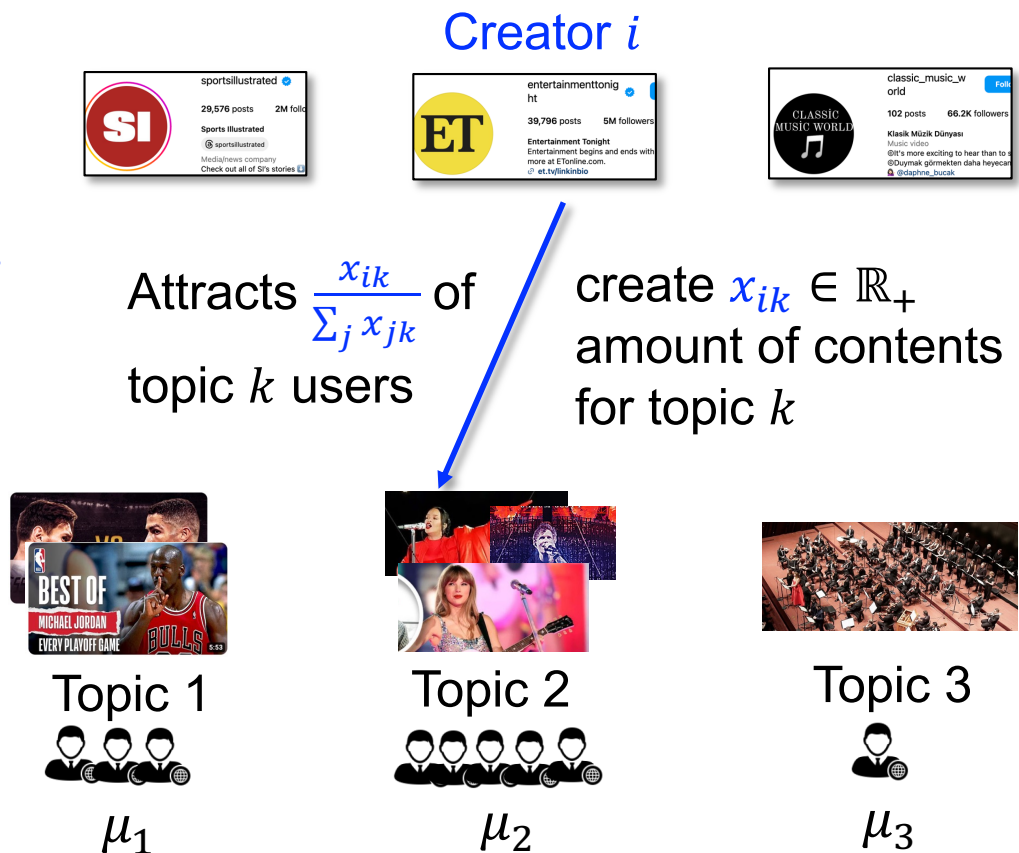


Similar studies of competition among (only) human creators has been widely studied recently, and also been applied to real systems

Payoff of creator $i =$

$$\sum_k \frac{x_{ik} \cdot \mu_k}{\sum_j x_{jk}} - c_i(\mathbf{x}_i)$$

- i 's cost of creating contents of different topics
- c_i captures i 's expertise (i.e., $c_i(\mathbf{x}) = x_1^2 + 0.1x_2^2$)

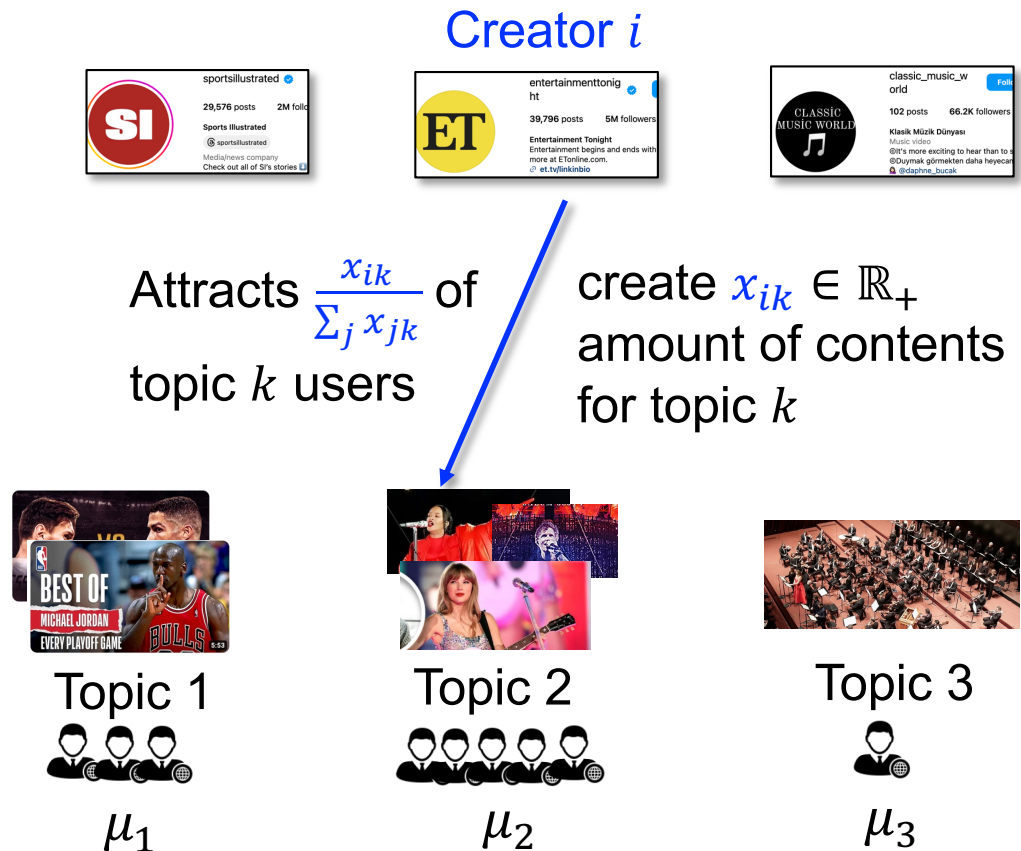


Similar studies of competition among (only) human creators has been widely studied recently, and also been applied to real systems

Next: integrate GenAI into the competition



- ✓ GenAI as a new way for content creation
- ✓ May affect creators in two different ways

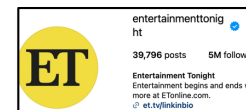


Model I: Exclusive Human-vs-GenAI Competition

- A special creator has exclusive access to GenAI (e.g., OpenAI or early adopters)
- Capture early stage of GenAI (e.g. now)

special GenAI creator

Creator i



Create $\alpha \cdot (\sum_j x_{jk})^\beta$
amount of contents

create $x_{ik} \in \mathbb{R}_+$
amount of contents
for topic k



Topic 1



μ_1



Topic 2



μ_2



Topic 3



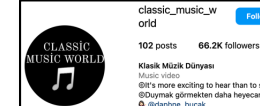
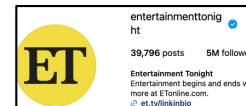
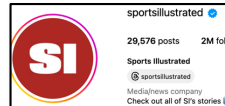
μ_3

Model I: Exclusive Human-vs-GenAI Competition

- A special creator has exclusive access to GenAI (e.g., OpenAI or early adopters)
- Capture early stage of GenAI (e.g. now)

special GenAI creator

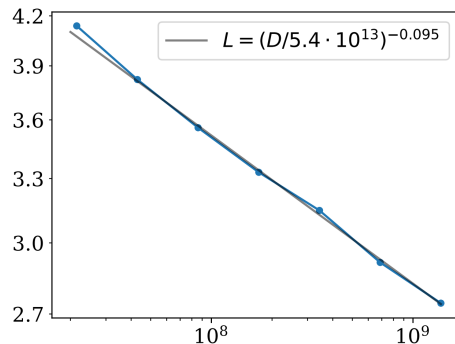
Creator i



α = GenAI data usage efficiency
 β = scaling exponent

Create $\alpha \cdot (\sum_j x_{jk})^\beta$
 amount of contents

create $x_{ik} \in \mathbb{R}_+$
 amount of contents for topic k



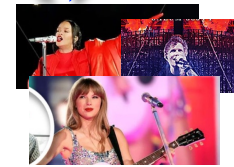
Dataset Size
 tokens
 [OpenAI, 2020]



Topic 1



μ_1



Topic 2



μ_2



Topic 3



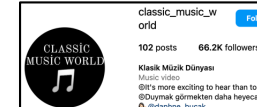
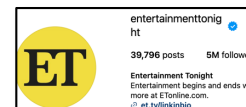
μ_3

Model I: Exclusive Human-vs-GenAI Competition

- The same model, except that the share of users attracted to creator i is now $\frac{x_{ik}}{\alpha \cdot (\sum_j x_{jk})^\beta + \sum_j x_{jk}}$ on topic k

special GenAI creator

Creator i



Create $\alpha \cdot (\sum_j x_{jk})^\beta$
amount of contents

create $x_{ik} \in \mathbb{R}_+$
amount of contents
for topic k

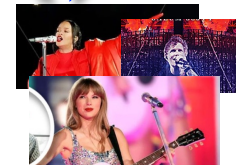
Competition captures the **tension** between GenAI and human creators



Topic 1



μ_1



Topic 2



μ_2



Topic 3



μ_3

Model II: Inclusive Human-vs-GenAI Competition

➤ Here GenAI technology has become mature enough that it is an option to every creator



a low-cost option that creates $\alpha \cdot \left(\sum_j x_{jk}\right)^\beta$ amount of contents

Creator i



create $x_{ik} \in \mathbb{R}_+$ amount of contents for topic k

This competition captures **tradeoff** between adopting GenAI and creating authentic contents



Topic 1



μ_1



Topic 2



μ_2



Topic 3



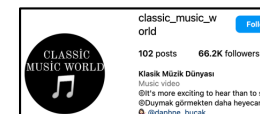
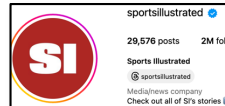
μ_3

Next:

- Some initial insights we obtained via game-theoretic analysis
- Will start from **exclusive** human-vs-GenAI competition

special GenAI creator

Creator i



Create $\alpha \cdot (\sum_j x_{jk})^\beta$
amount of contents

create $x_{ik} \in \mathbb{R}_+$
amount of contents
for topic k



Topic 1



μ_1



Topic 2



μ_2



Topic 3



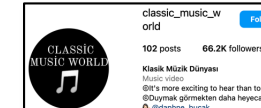
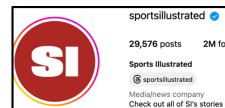
μ_3

Exclusive Human-vs-GenAI Competition

Q1: does a (pure) Nash equilibrium (x_1^*, \dots, x_n^*) always exist?

special GenAI creator

Creator i



Create $\alpha \cdot (\sum_j x_{jk})^\beta$
amount of contents

create $x_{ik} \in \mathbb{R}_+$
amount of contents
for topic k



Topic 1



μ_1



Topic 2



μ_2



Topic 3



μ_3

Exclusive Human-vs-GenAI Competition

Q1: does a (pure) Nash equilibrium (x_1^*, \dots, x_n^*) always exist?

- Fortunately, YES if $\beta \in [0,1]$ and c_i convex – in fact, it's unique
 - This is a pleasant surprise, since equilibrium of Tullock contests is generally complex
- The new $\alpha \cdot (\sum_j x_{jk})^\beta$ term does not show up in any classic study of Tullock contests
 - A unique feature of ML that depends on aggregated data



Create $\alpha \cdot (\sum_j x_{jk})^\beta$
amount of contents

$\beta =$ scaling exponent

Exclusive Human-vs-GenAI Competition

Q1: does a (pure) Nash equilibrium (x_1^*, \dots, x_n^*) always exist?

- Key proof idea: it can be shown to be a monotone game
 - Side product: many natural no-regret learning dynamics converge to equilibrium
- Moreover, this equilibrium provably admits natural properties:
 - More cost-efficient human creators generate more contents
 - More human creators lead to more total body of contents

Exclusive Human-vs-GenAI Competition

Q2: Will GenAI drive humans out of the ecosystems? Are there still authentic contents at equilibrium and, if so, how much?

➤ A fortunately positive answer – they will reach certain “symbiosis”

Theorem [YLNWX'24]: Suppose cost function is *weighted* l_ρ -norm (i.e., $c_i(\mathbf{x}_i) = \sum_k c_{ik} (x_{ik})^\rho$), then at the unique pure NE we have

$$\frac{(s_k^*)^\rho}{\mu_k \left\| \mathbf{c}_{\cdot k}^{-1} \right\|_{\frac{1}{\rho-1}}} = C(\alpha, \rho)$$

where $s_k^* = \sum_i x_{ik}^*$ is the total human-created content for topic k and $\mathbf{c}_{\cdot k}^{-1} = (c_{1k}^{-1}, c_{2k}^{-1}, \dots, c_{nk}^{-1})$ captures human production efficiency.

Exclusive Human-vs-GenAI Competition

Theorem [YLNWX'24]: Suppose cost function is weighted l_ρ -norm (i.e., $c_i(\mathbf{x}_i) = \sum_k c_{ik} (x_{ik})^\rho$), then at the unique pure NE we have

$$\frac{(s_k^*)^\rho}{\mu_k \left\| \mathbf{c}_{\cdot k}^{-1} \right\|_{\frac{1}{\rho-1}}} = C(\alpha, \rho)$$

where $s_k^* = \sum_i x_{ik}^*$ is the total human-created content for topic k and $\mathbf{c}_{\cdot k}^{-1} = (c_{1k}^{-1}, c_{2k}^{-1}, \dots, c_{nk}^{-1})$ captures human production efficiency.

- GenAI's learning rate β did not show up in the bound (though it did affect equilibrium existence)
- Note $\alpha = 0$ degenerates to no GenAI case, so competition with GenAI will not drive human out, but reduce total creation by a constant factor
- Further analysis shows that as human become more efficient (i.e., $\downarrow \rho$),
 - more contents will be created for niche topic (small μ_k)
 - less contents will be created for popular topic (large μ_k)

Exclusive Human-vs-GenAI Competition

Theorem [YLNWX'24]: Suppose cost function is weighted l_ρ -norm (i.e., $c_i(\mathbf{x}_i) = \sum_k c_{ik} (x_{ik})^\rho$), then at the unique pure NE we have

$$\frac{(s_k^*)^\rho}{\mu_k \left\| \mathbf{c}_{\cdot k}^{-1} \right\|_{\frac{1}{\rho-1}}} = C(\alpha, \rho)$$

where $s_k^* = \sum_i x_{ik}^*$ is the total human-created content for topic k and $\mathbf{c}_{\cdot k}^{-1} = (c_{1k}^{-1}, c_{2k}^{-1}, \dots, c_{nk}^{-1})$ captures human production efficiency.

Remarks.

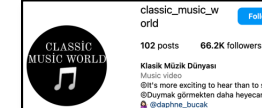
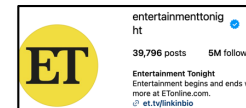
- Empirical simulation reveals that similar insights hold for general cost functions
- All above results generalize to the $\mu_k(s_k)$ cases

Inclusive Human-vs-GenAI Competition

a low-cost option that creates $\alpha \cdot (\sum_j x_{jk})^\beta$ amount of contents



Creator i



create $x_{ik} \in \mathbb{R}_+$ amount of contents for topic k



Topic 1



μ_1



Topic 2



μ_2



Topic 3



μ_3

Inclusive Human-vs-GenAI Competition

Q1: does a (pure) Nash equilibrium always exist?

- Unfortunately, not always, but YES under certain conditions
 - Existence requires “**light competition**” – i.e., sufficiently good scaling exponent of GenAI and fast-growing user demand
 - Generally, not unique

Inclusive Human-vs-GenAI Competition

Q2: How many human creators will resort to GenAI for content creation, and who are they?

Theorem [YLNWX'24]: Assume l_ρ -norm costs and light competition, then there always exists a pure Nash equilibrium of following format:

on each topic k , creation at equilibrium is $(x_{1,k}, x_{2,k}, \dots, x_{m,k}, \perp, \dots, \perp)$ where $\perp = \text{GenAI}$ and $c_{1,k} \leq c_{2,k} \leq \dots \leq c_{m,k} < c_{m+1,k} \leq \dots \leq c_{n,k}$.

Moreover,

$$\frac{m}{n} \leq C(\alpha, \beta, \rho) \cdot \frac{(\mu_k)^{\frac{1-\beta}{\rho}}}{n^{1-\frac{(1-\beta)(\rho-1)}{\rho}}}$$

Takeaways

- Unsurprisingly, less efficient creators will switch to GenAIs
- Less fraction of creators generate authentic contents when (a) topic is niche (small μ_k); or (b) competitive environment (large n); or (c) better GenAI technology
- When n is significantly larger than μ_k , everyone switches to GenAI

Inclusive Human-vs-GenAI Competition

Q2: How many human creators will resort to GenAI for content creation, and who are they?

Theorem [YLNWX'24]: Assume l_ρ -norm costs and light competition, then there always exists a pure Nash equilibrium of following format:

on each topic k , creation at equilibrium is $(x_{1,k}, x_{2,k}, \dots, x_{m,k}, \perp, \dots, \perp)$ where $\perp = \text{GenAI}$ and $c_{1,k} \leq c_{2,k} \leq \dots \leq c_{m,k} < c_{m+1,k} \leq \dots \leq c_{n,k}$.

Moreover,

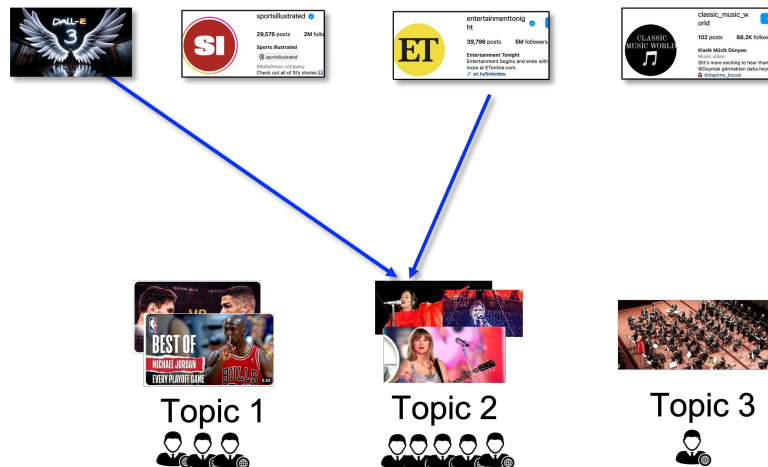
$$\frac{m}{n} \leq C(\alpha, \beta, \rho) \cdot \frac{(\mu_k)^{\frac{1-\beta}{\rho}}}{n^{1-\frac{(1-\beta)(\rho-1)}{\rho}}}$$

Remarks

- These insights are also observed in simulations, for more general setups
- All above results generalize to the $\mu_k(s_k)$ cases

Summary

- Game-theoretic analysis about the competition of GenAI and human for content creation in recommendation systems
- Two different modes of competition, depending on the stage of the AI technology
- Economic analysis shows encouraging symbiosis between human and GenAI
 - No significant harm on the macro-level, but does not rule out possibility that a single individual human can do significantly worse

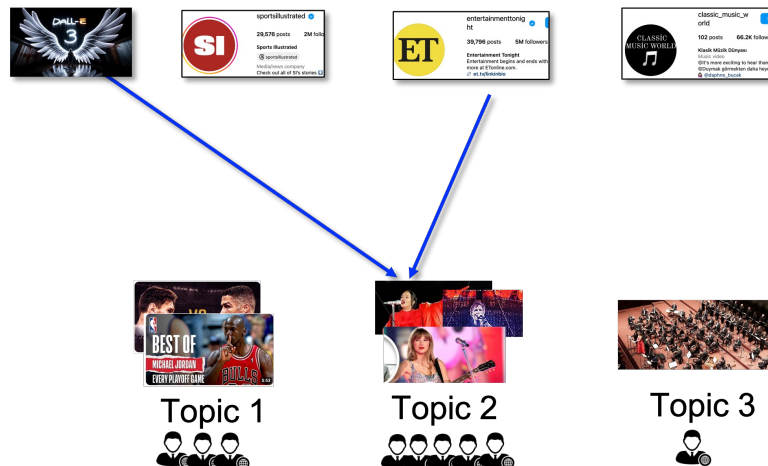


Many Many Open Directions



Credit: DALL.E

- Copyright issue
- What if GenAI platforms have to pay creators for acquiring their authentic data for training GenAIs, and how to acquire such data?
- What if GenAI platforms charge creators for generating contents?
- Authentic contents can now profit from serving Internet users **AND** serving GenAI platforms – how would this change the competition?

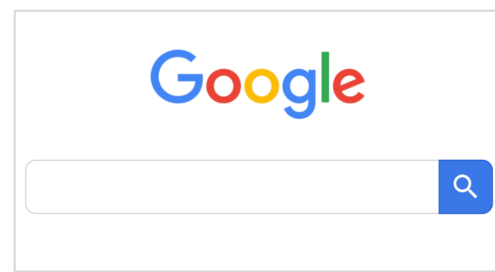
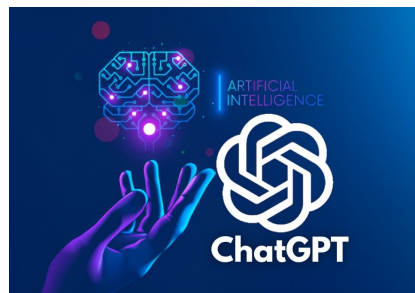


Many Many Open Directions



Credit: DALL.E

- Copyright issue
- What if GenAI platforms have to pay creators for acquiring their authentic data for training GenAIs, and how to acquire such data?
- What if GenAI platforms charge creators for generating contents?
- Authentic contents can now profit from serving Internet users **AND** serving GenAI platforms – how would this change the competition?
- What about the “search engine + conversational engine” competition?



End of Lecturing for CMSC 3540 I

Hope you enjoyed the topics!

Next: Your Show Time



Thank You

Questions?

haifengxu@uchicago.edu