

Conversations in LLM Applications

Marti Hearst
UC Berkeley

In collaboration with the Allen Institute of AI



MIDS Immersion

November 7, 2023

Talk Outline

01

Generative AI / 'Natural' UIs

02

Designing for Conversation

03

Designing for Automation Bias

01

Generative AI & ‘Natural’ UIs

Generative AI is Hugely Disruptive

GENERATIVE AI: BY THE NUMBERS	
8	Estimated value in billions of US dollars of the global GenAI market in 2023. ²
65	Months it took Twitter to reach 100 million users after launch. ³
2	Months it took ChatGPT to achieve that milestone. ²
2	Rank of ChatGPT among all apps that have ever reached that mark. ²
43	Estimated percentage of college students who have used ChatGPT or a similar app. ⁴
80	Estimated percentage of U.S. workers who will have at least 10% of their work tasks affected by GenAI. ⁵
19	Percentage of such workers who will have at least 50% of their tasks so affected. ⁴
3.4	Gigawatt hours of electricity used to train Google's PaLM for two months. ⁶
321	Number of U.S. homes whose needs could be fully met with that power for a full year. ⁷
59	Millions of total words used in the 20-volume <i>Oxford English Dictionary</i> (2nd ed.). ⁹
1,400,000	Millions of words (or word parts) in the dataset used to train Meta AI's Llama. ⁸
76	Estimated percentage of U.S. consumers concerned about GenAI-produced misinformation. ¹⁰

ILLUSTRATION: ©IAREMENKO

Generative AI is Hugely Disruptive

43	Estimated percentage of college students who have used ChatGPT or a similar app. ⁴
80	Estimated percentage of U.S. workers who will have at least 10% of their work tasks affected by GenAI. ⁵
19	Percentage of such workers who will have at least 50% of their tasks so affected. ⁴

Generative AI is Hugely Popular

65	Months it took Twitter to reach 100 million users after launch. ³
2	Months it took ChatGPT to achieve that milestone. ²

Generative AI is Hugely Popular

 **TheUpshot**

35 Ways Real People Are Using A.I. Right Now

By Francesca Paris and Larry Buchanan April 14, 2023

35 Ways Real People Are Using A.I. Right Now

By Francesca Paris and Larry Buchanan April 14, 2023

35 Ways Real People Are Using A.I. Right Now

By Francesca Paris and Larry Buchanan April 14, 2023

Play devil's advocate
(to anticipate counter-arguments)

Design a new game

Transcribe clinical notes

Comment on the wording of an assignment

Plan Meals

Plan Workouts

Ask questions about papers

Plan Gardens

Correct dictated text

Organize a computer directory

Write a wedding speech

Make a spotify playlist

Write excel formulas

Start filling the “blank page”

Find and aggregate scholarly research

Why is ChatGPT so popular?

**A major reason: natural
language as the input**

Why is ChatGPT so popular?

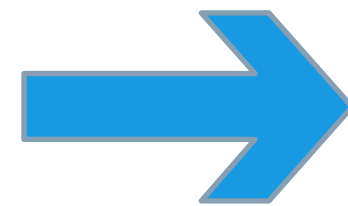
**A major reason: natural
language as the input**

**In general, I/O is better conforming to
how people think, communicate, move**

Gradually moving to 'Natural' User Interfaces

‘Natural’ User Interfaces

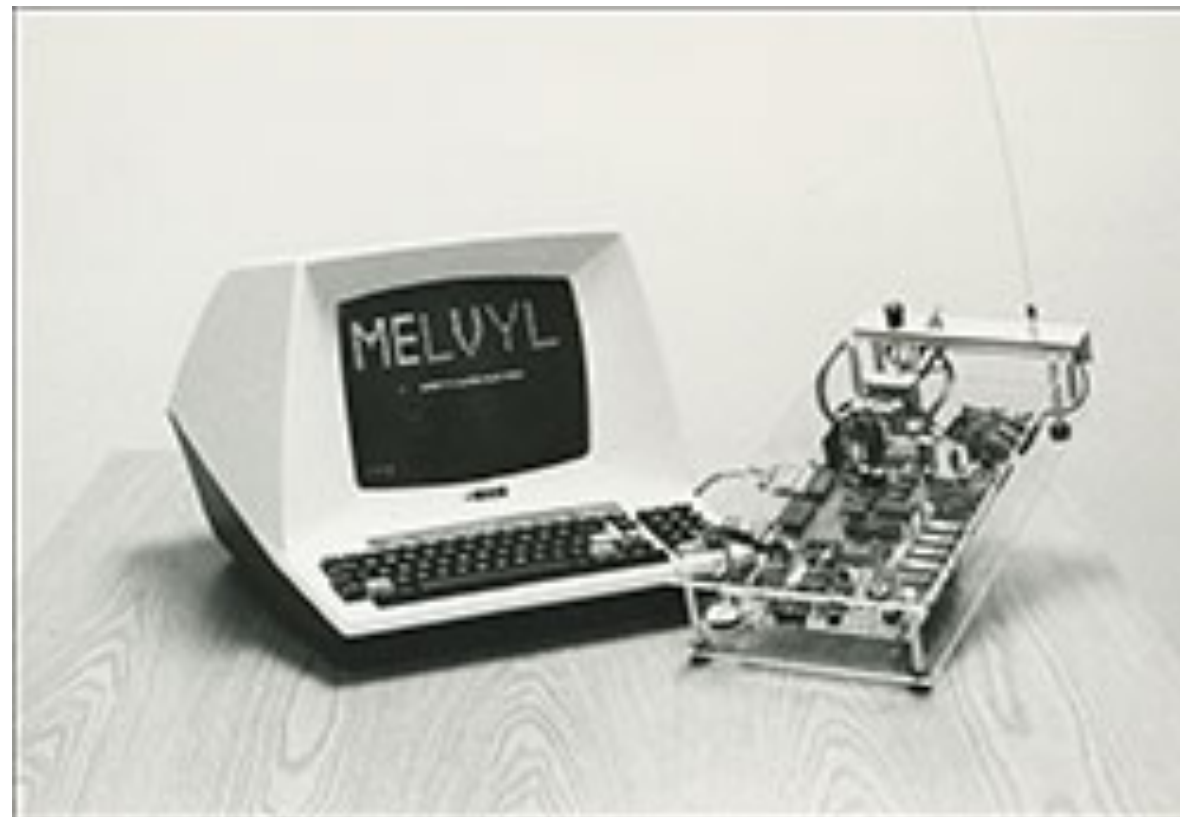
Part of a larger trend



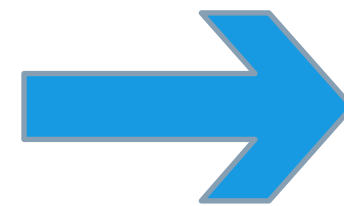
Accessories sold separately

‘Natural’ User Interfaces

Part of a larger trend



```
0. College and research libraries  
{Chicago} : American Library Association, [1979]  
Also issued online  
Vol. for 1966 has supplements A(21) and B(21). For 1967-1991  
have supplements College and research libraries news  
  
Web Resources: No call number  
Library has: V.5(1960-4(1990-2013))-WilsonWeb Library L15.  
Restricted to UC campuses  
Electronic locations:  
http://ogonvul.cdlib.org/  
7950-DCP1DCP4grade=article4__dual_mes=41E04588A-0010-0070  
  
Curr Per: 1471.C6  
Not-circulating.  
Library has: Current Issues  
  
Main Stack: 1471.C6  
Shelved: Current unbound issues in Current Periodicals.  
Library has: BOUND 1(1919)-60(2007)-  
Library has: CUMULATIVE INDEXES 1(1919)-11(1954),  
21(1940)-50(1999)  
Shelved at NRLF: v.1-34  
v.1-2 (Dec 1919-Sept 1941) NRLF #1 D 4 143 C46 BAKO00E1  
C021161194  
v.3-4 (Dec 1941-Sept 1943) NRLF #1 D 4 143 C47 BAKO00E1  
C021161100  
v.5-6 (Dec 1943-Sept 1945) NRLF #1 D 4 143 C48 BAKO00E1  
C021161119  
v.7-8 (1946-47) NRLF #1 D 4 143 C49 BAKO00E1 C021161120
```



Search History
[Database Details](#)

Set	Term Searched	Items	
S1	MELANOMA OR SKIN (3N)CANCER?	1264	Display
S2	PREVENT? OR PROTECT?	28430	Display
S3	ULTRAVIOLET OR UV OR SUN OR SUNSCREEN	1886	Display

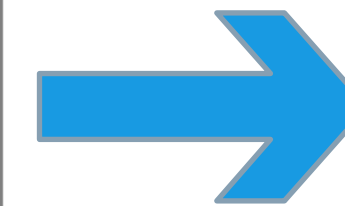
Format: Number of Records:

Show Database Details for:
 [Bluesheet](#) [Rates](#)

[Fields](#) [Formats](#) [Sorts](#) [Limits](#) [Tags](#)

© 1998 The Dialog Corporation, Inc.

Command [Submit](#) [Previous](#)



Berkeley Library
NEW SEARCH JOURNAL SEARCH DATABASES BROWSE SEARCH ...

UC LIBRARY SEARCH automation bias Articles, books, and more ADVANCED SEARCH

Refine your results

Sort by Relevance

Show Only

Available online

Peer-reviewed Journals

Open Access

Held by UC

Resource Type

Author/Creator

Subject

Language

Date

Library

UC Libraries

Genre

Journal Title

0 selected PAGE 1 1-10 of 14,971 Results Save query

1 BOOK Unpacking Skill Bias: Automation and New Tasks Acemoglu, Daron; National Bureau of Economic Research; Restrepo, Pascual. 2020 Available Online

2 BOOK Career Exploration in Computers Project. Final Report. Hazard Community Coll., Ky. 1987 Available at Northern Regional Library Facility NRLF Shared Print (FedDocArc) (ED344022)

3 ARTICLE Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice Alon-Barkat, Saar; Busuioc, Madalina Journal of public administration research and theory, 2023, Vol.33 (1), p.153-169, Article 153 44 ...: overreliance on algorithmic advice even in the face of "warning signals" from other sources (automation bias), and selective adoption of algorithmic advice when this corresponds to stereotype... 44 PEER REVIEWED OPEN ACCESS Available Online

4 ARTICLE Mimes of automated decision aids: Consequences, automation bias, and

- UI went from CRT commands to bitmap commands to modern UI
- Librarians encourage patrons to have a detailed conversation
- Converted to category labels, and Boolean search over surrogates of known sources

Web search:

An earlier information access revolution

- 20 years ago: People all backgrounds using keyword search and looking through links for answers
- This was very foreign!
- Many people disliked this paradigm

SearchPad

AltaVistaConnections [AltaVista.com](#) [My AltaVista](#)

Ask AltaVista™ a question. Or enter a few words in

Search For: ☒ Web Pages ☐ Images ☐ Video ☐ Audio

Related Searches:

- [uses of genetic engineering](#)
- [Genetic Engineering News](#)
- [Human genetic engineering](#)
- [Plant Genetic Engineering](#)
- [Genetic Engineering Network](#)
- [genetic engineering](#)
- [Ethics of Genetic Engineering](#)
- [What is genetic engineering](#)

Featured Sites

Shop at [Amazon.com](#)

- [Books: genetic engineering](#)
- [CD, Video, Auctions](#)

About.com

- [Find: genetic engineering](#)
- [Send a card!](#)

Shopping.com

- [Search: genetic engineering](#)

AltaVista Shopping

- [Visit our online shopping guide](#)

[Click Here!](#)

► **AltaVista found 52839 Web pages.**

GENETIC

The RealNames link takes you directly to **GENETIC**.

Mark 1. What is genetic engineering?

What is genetic engineering? Genetic engineering is the term used to describe techniques in molecular biology that have literally...

URL: [www.aba.asn.au/leaf2.html](#)

Last modified 18-Jun-96 - page size 10K - in English [[Translate](#)]

Mark 2. Genetic Engineering and Its Dangers

For an updated version of this page, click here. Please change your browser to Netscape 4.0 or higher. Compiled by Dr Ron Epstein. Philosopher

URL: [userwww.sfsu.edu/~rone/Genetic%20Engineering.htm](#)

Last modified 21-Mar-97 - page size 7K - in English [[Translate](#)]

Without good language analysis, there are many challenges in designing search UIs

How to disambiguate queries?

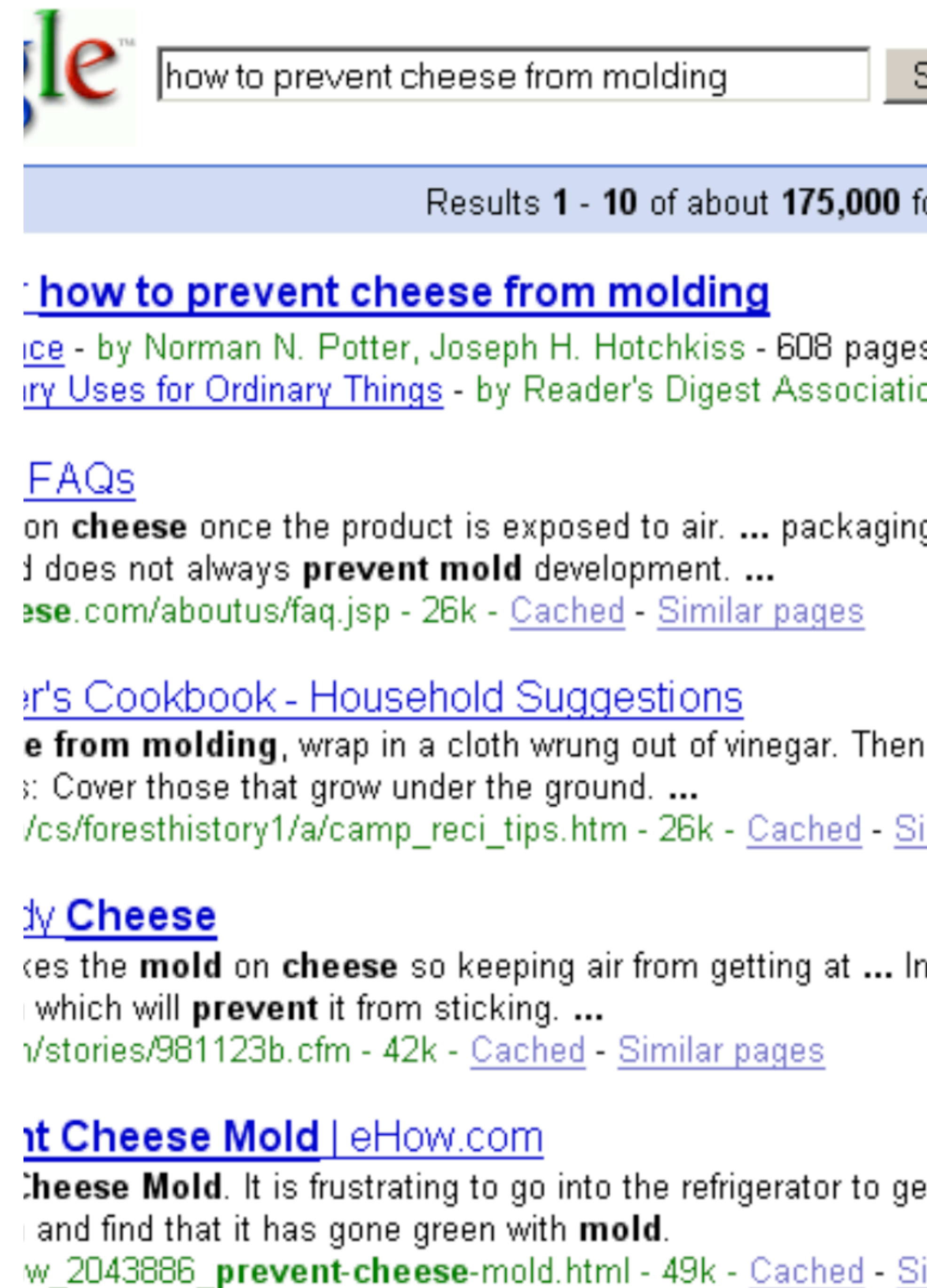
How to design the “snippet”

How to rank full text given a few keywords?

How to respond to short queries?

How to organize thousands of results?

How to suggest alternative queries?



The screenshot shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "how to prevent cheese from molding". To the right of the search bar is a button with the letter "S". Below the search bar is a blue horizontal bar with the text "Results 1 - 10 of about 175,000 for". Below this bar, the first search result is displayed in blue text: "[how to prevent cheese from molding](#)". Below the title is a snippet of text in green: "ice - by Norman N. Potter, Joseph H. Hotchkiss - 608 pages" and "ry Uses for Ordinary Things - by Reader's Digest Associati". Below the snippet is a link in blue: "FAQs". Below the link is another snippet of text in green: "on **cheese** once the product is exposed to air. ... packaging" and "d does not always **prevent mold** development. ...". Below the snippet is a link in green: "ese.com/aboutus/faq.jsp - 26k - [Cached](#) - [Similar pages](#)". Below the link is another search result in blue text: "[er's Cookbook - Household Suggestions](#)". Below the title is a snippet of text in green: "e **from molding**, wrap in a cloth wrung out of vinegar. Then" and "s: Cover those that grow under the ground. ...". Below the snippet is a link in green: "/cs/foresthstory1/a/camp_recip_tips.htm - 26k - [Cached](#) - [Si](#)". Below the link is another search result in blue text: "[dy Cheese](#)". Below the title is a snippet of text in green: "ces the **mold** on **cheese** so keeping air from getting at ... In" and "which will **prevent** it from sticking. ...". Below the snippet is a link in green: "v/stories/981123b.cfm - 42k - [Cached](#) - [Similar pages](#)". Below the link is another search result in blue text: "[nt Cheese Mold | eHow.com](#)". Below the title is a snippet of text in green: "heese **Mold**. It is frustrating to go into the refrigerator to ge" and "and find that it has gone green with **mold**". Below the snippet is a link in green: "w_2043886_prevent-cheese-mold.html - 49k - [Cached](#) - [Si](#)".

Natural Language as a UI for Search



what is the height of a typical giraffe



All

Images

Shopping

News

Videos

More

Tools

Giraffe / Height

16 – 19 ft.

Male, Adult

14 – 17 ft.

Female, Adult

<https://zoologist.com> › how-tall-is-a-giraffe

[How Tall is a Giraffe - Giraffe Height - Zoologist](https://zoologist.com/how-tall-is-a-giraffe/)

Apr 23, 2019 — The average height of a Giraffe in feet is **16.4 to 19.7 feet**. 2. How Tall is a Male Giraffe. The average height of a male Giraffe is 5 to 6 ...

<https://www.pbs.org> › Home › Blog

43

[Giraffe Fact Sheet | Blog | Nature - PBS](https://www.pbs.org/blog/giraffe-fact-sheet/)

Jul 15, 2020 — Size and Weight: ... Giraffes are the tallest land animals. Female giraffes are **up to 14 feet tall** and weigh up to 1,500 pounds. Meanwhile, males ...

<https://animals.mom.com> › tall-giraffes-1592

[How Tall are Giraffes? - Pets on Mom.com](https://animals.mom.com/tall-giraffes-1592/)



Giraffe

Animal



The giraffe is a large African hoofed mammal

With conversations that really work, no need to worry about those earlier design concerns



what is the height of a typical giraffe?



[View other drafts](#)



The typical height of a giraffe is 14-18 feet (4.3-5.5 meters) tall, with males being taller than females. The tallest giraffe ever recorded was 20 feet (6.1 meters) tall.

Giraffes are the tallest land mammals in the world, and their height gives them a number of advantages. For example, it allows them to reach leaves and other food sources that are out of reach for other animals. It also gives them a good vantage point for spotting predators and other threats.

Giraffes' long necks are also an important adaptation to their height. Their necks contain seven vertebrae, the same number as humans, but each vertebra is much longer. This allows giraffes to reach leaves and other food sources that are high up without having to strain.

Giraffes' hearts are also specially adapted to their height. Their hearts are two feet long and weigh about 25 pounds. This allows them to pump blood to their heads even when they are standing up straight.

Giraffes are truly amazing creatures, and their height is one of their most distinctive features.

Natural Language as the UI for Search



People always loved Ask Jeeves even though it never actually answered questions

[MSN Encarta - Giraffe](#)

...grow to a **height** of almost 6 m (20 ft)â€”tall enough to span two floors of a **typical** office ...
In addition to its great **height**, th...

[encarta.msn.com/encyclopedia_761561060/Giraffe.html](#)

[Giraffe](#)

Tallest of all mammals, the **giraffe** attains an overall **height** of 5.5 meters (18 feet) or more. ...
contains only the seven vertebrae **typical** of

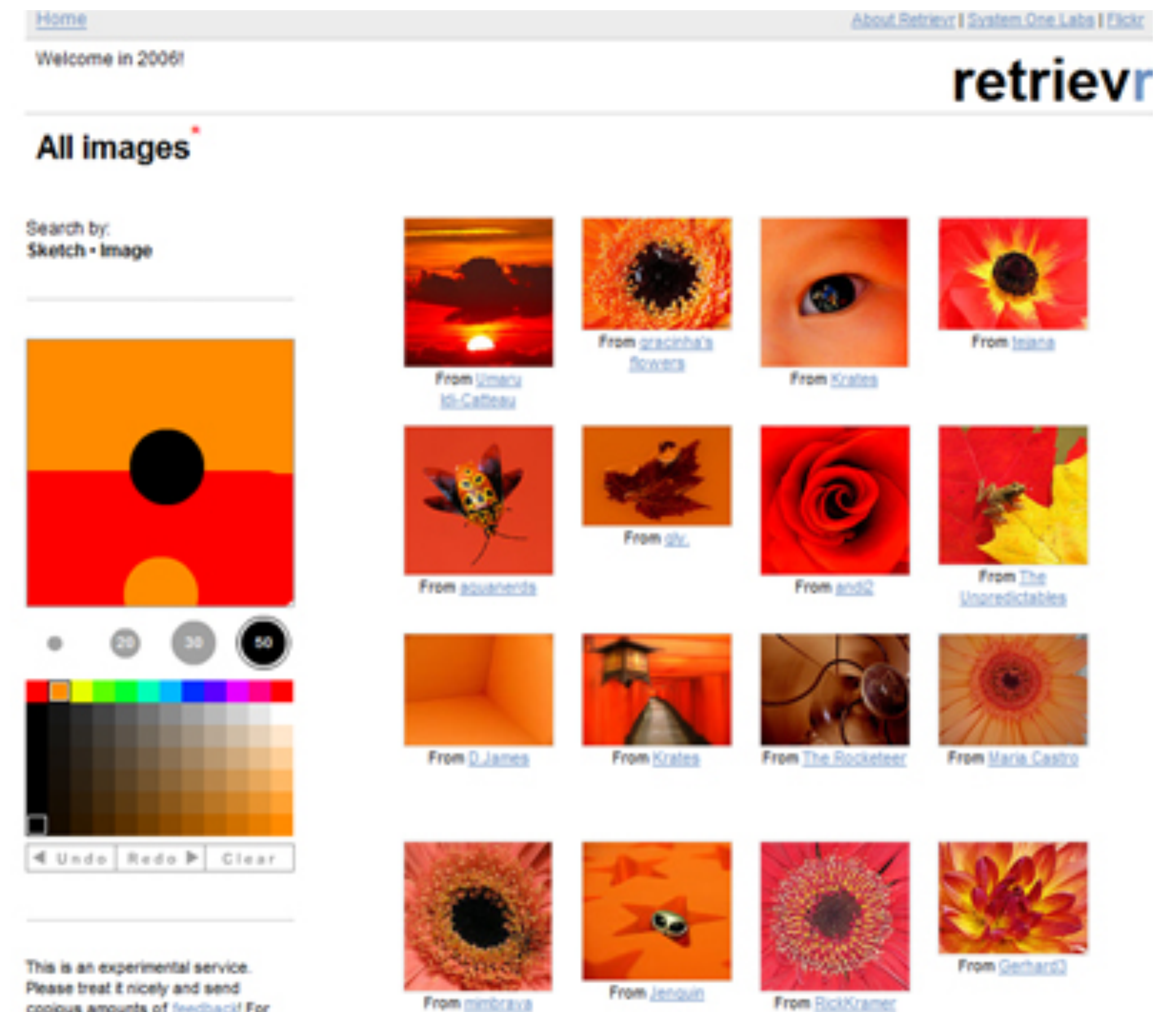
Similar story for image search

Lacking robust image analysis, images had to be queried by color and shape, or sketching

This was not ‘natural’ and did not work well.



Example taken from Jacobs, Finkelstein, & Salesin
Fast Multi-Resolution Image Querying, SIGGRAPH 1995



(Although Dall-E doesn't always generate what you want either!)

what are some good towns
to visit in Turkey?

Some charming towns to
explore in Turkey include
Cappadocia renowned for its
unique fairy chimney
formations, Ephesus for its
well-preserved ruins of the
Roman empire, and
Pamukkale for its famous

Language-based Chat is the new User Interface

tell me more about
Cappadocia

Cappadocia is a semi-arid
region in central Turkey
known for its distinctive fairy
chimneys, tall cone-shaped
rock formations clustered in
Monks Valley and elsewhere.

Send a message



But ... it has to be very conversation-like

- Studies show that people on average experiment with Alexa for about a week and then settle on a small fixed set of commands
- The technology was not ready to become a ‘natural’ UI



Consequence of 'Natural' User Interfaces

Less intuitive UIs exist because the technology isn't ready to make them natural

The more UIs conform to how people think, move, communicate, the broader the uptake and thus the more disruptive

Talk Outline

01

Generative AI / 'Natural' UIs

02

Designing for Conversations

03

Designing for Automation Bias

02

Designing for Conversations

Conversations

Conversations

are **interactive**

Human Conversations are Complex

Much effort is spent in maintaining & updating mutual understanding

Action	Amount (%)
Evaluating and comparing options	25%
Suggesting courses of action	23%
Clarifying and establishing state	13.5%
Clarifying or confirming the communication	13.5%
Discussing problem-solving strategy	10%
Summarizing courses of action	8%
Identifying problems and alternatives	7%

Designing for Conversations

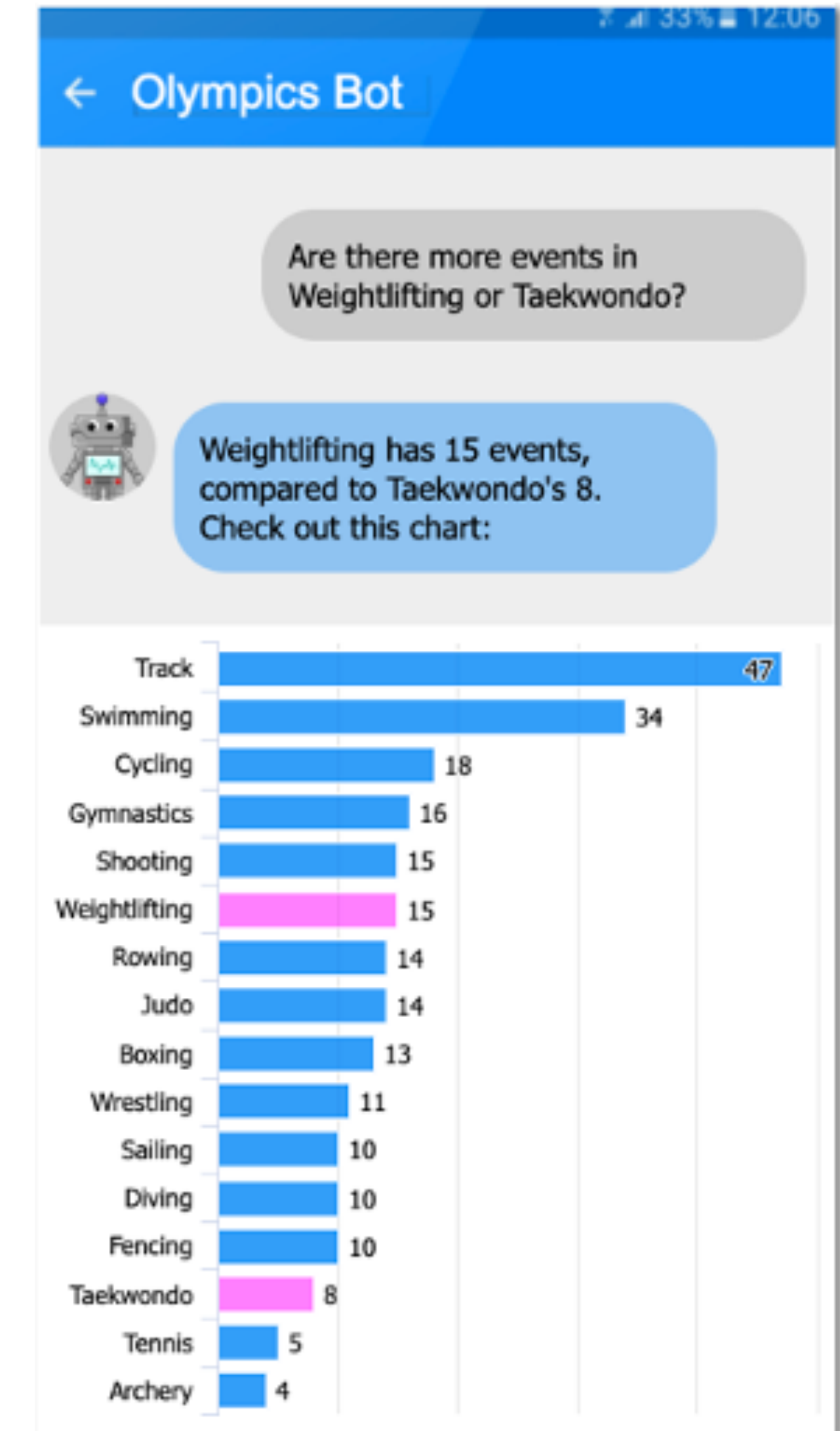
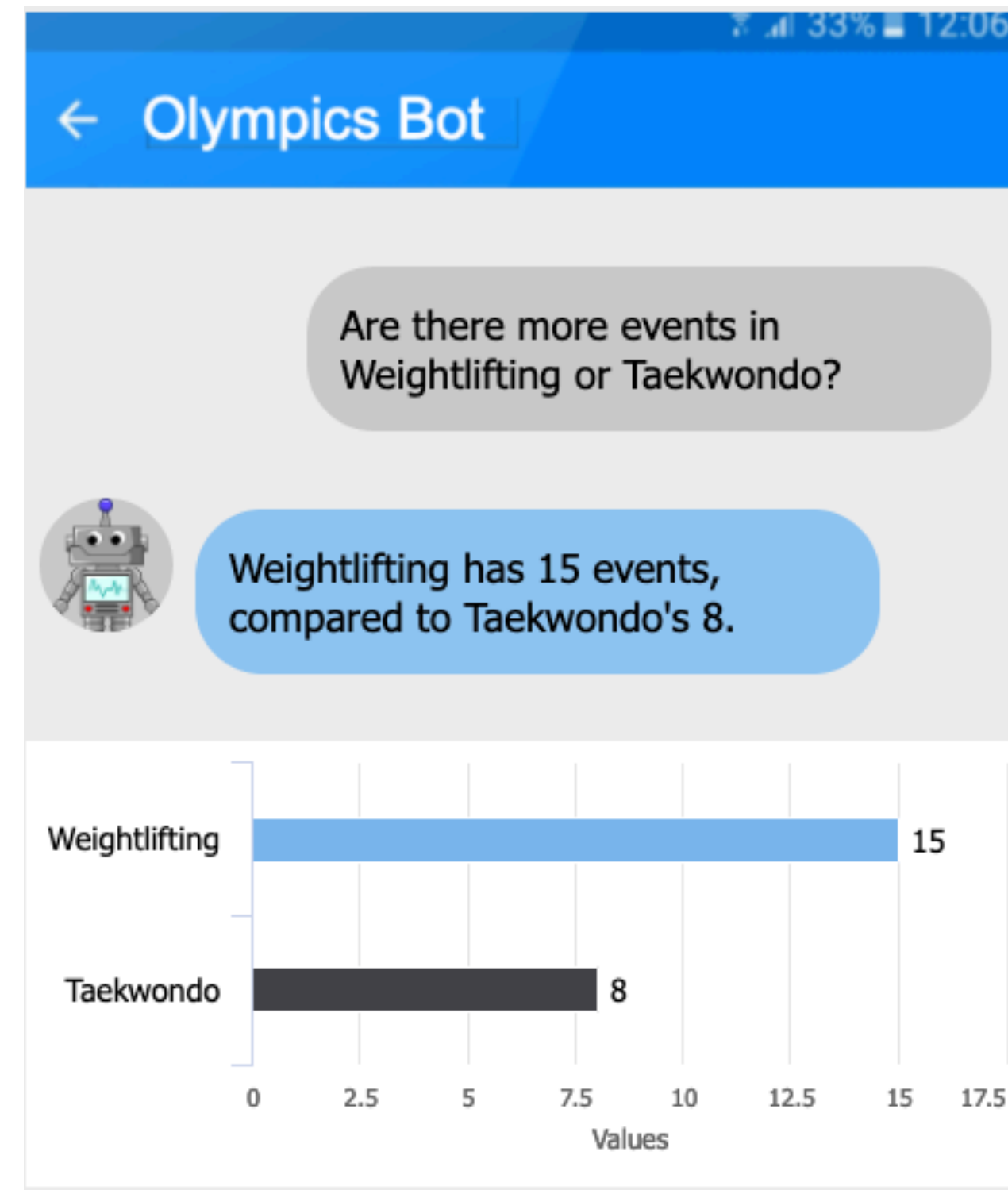
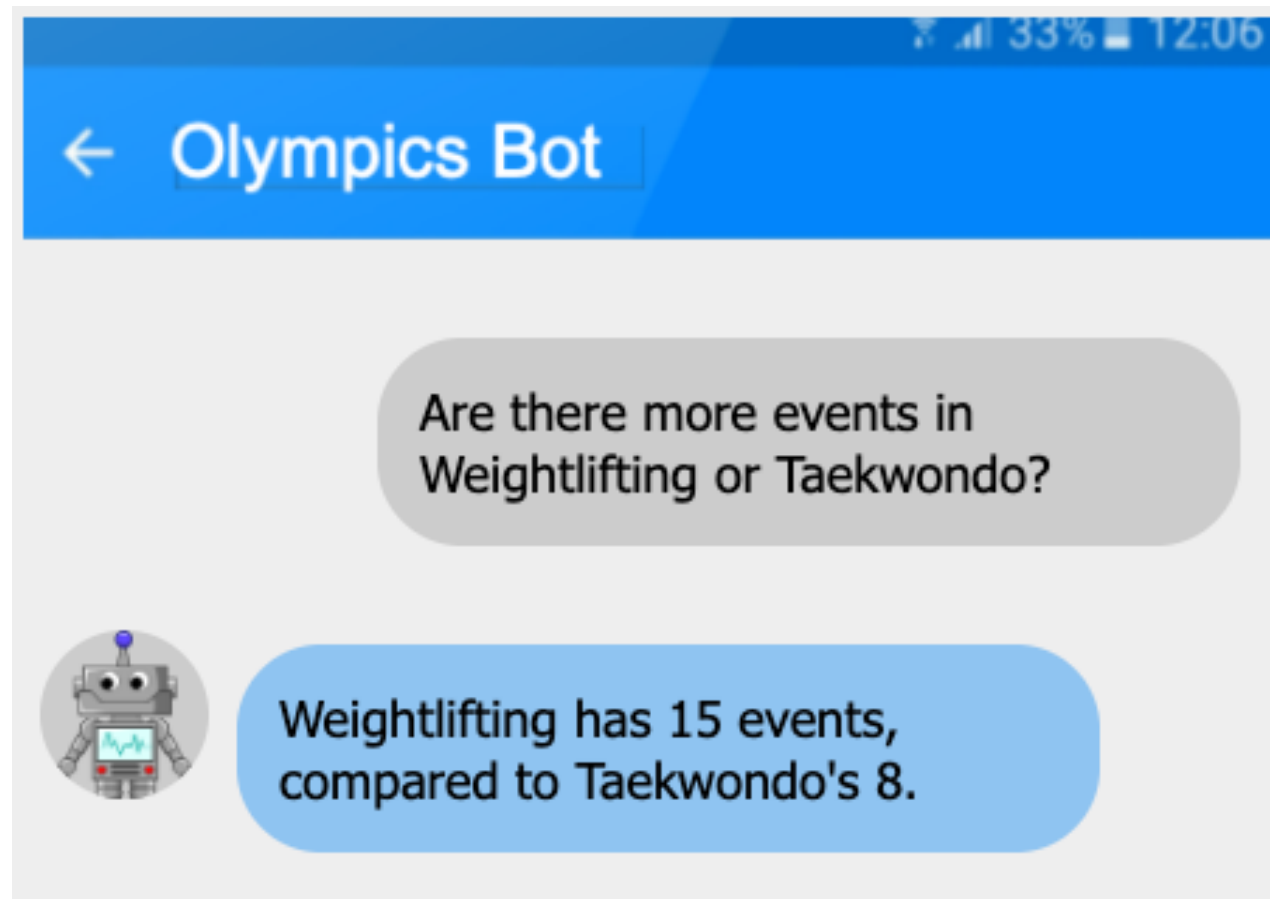
**Example: Do people want to see a
visualization in a chat conversation?**

Would You Like A Chart With That?

Incorporating Visualizations into Conversational Interfaces

Marti Hearst*

Melanie Tory†



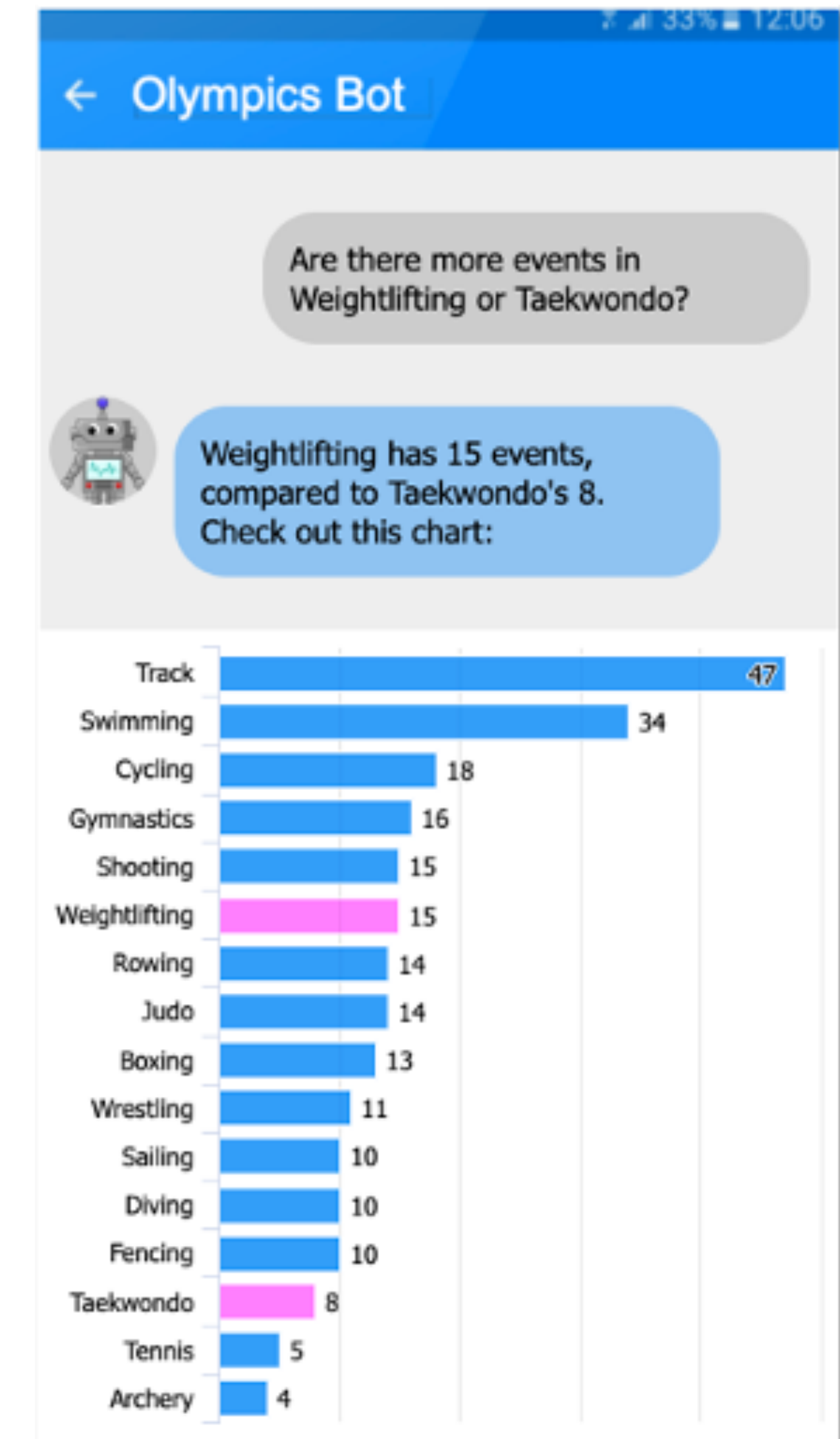
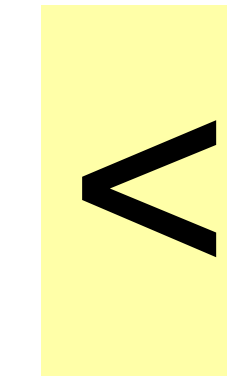
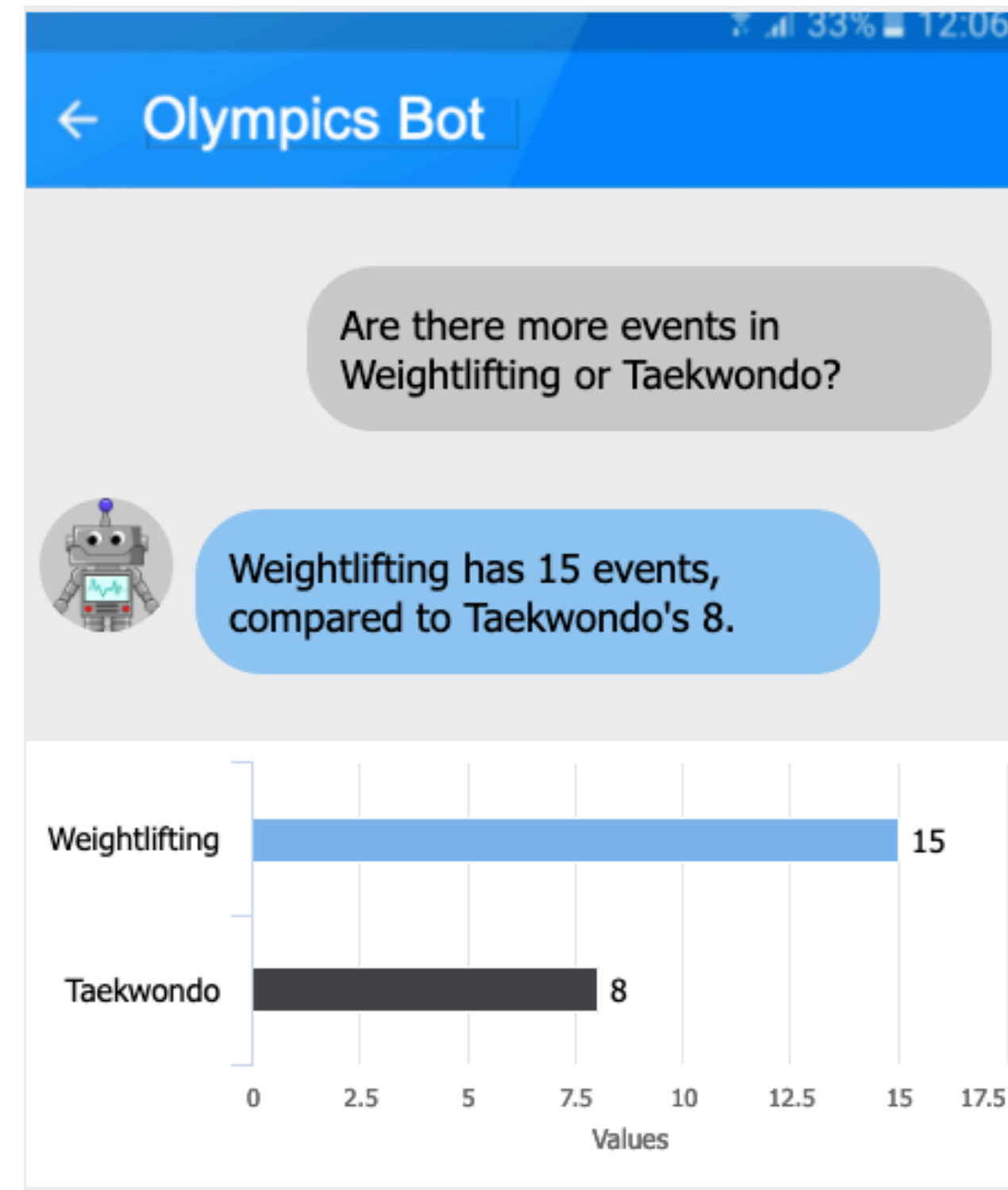
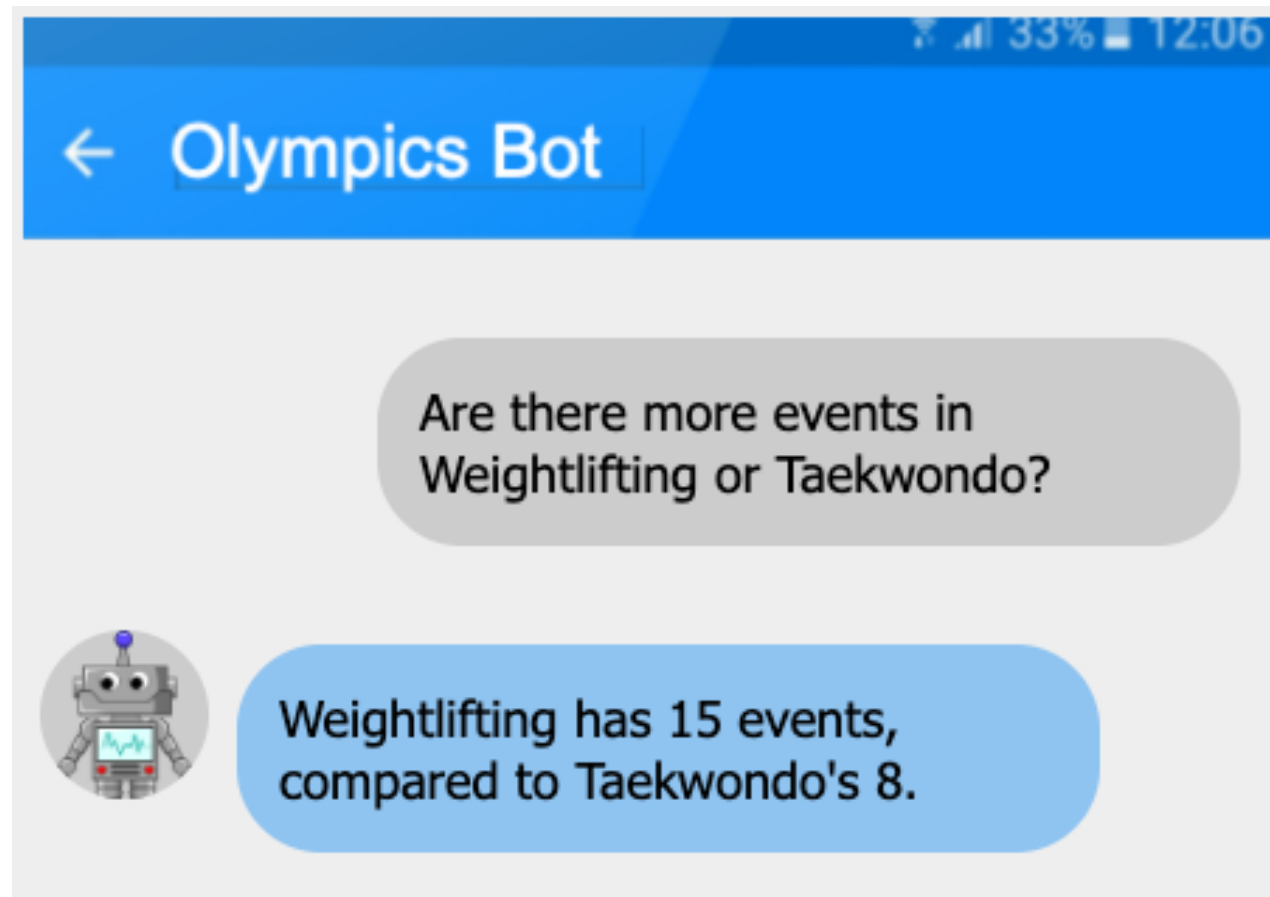
Which is preferred?

Would You Like A Chart With That?

Incorporating Visualizations into Conversational Interfaces

Marti Hearst*

Melanie Tory†

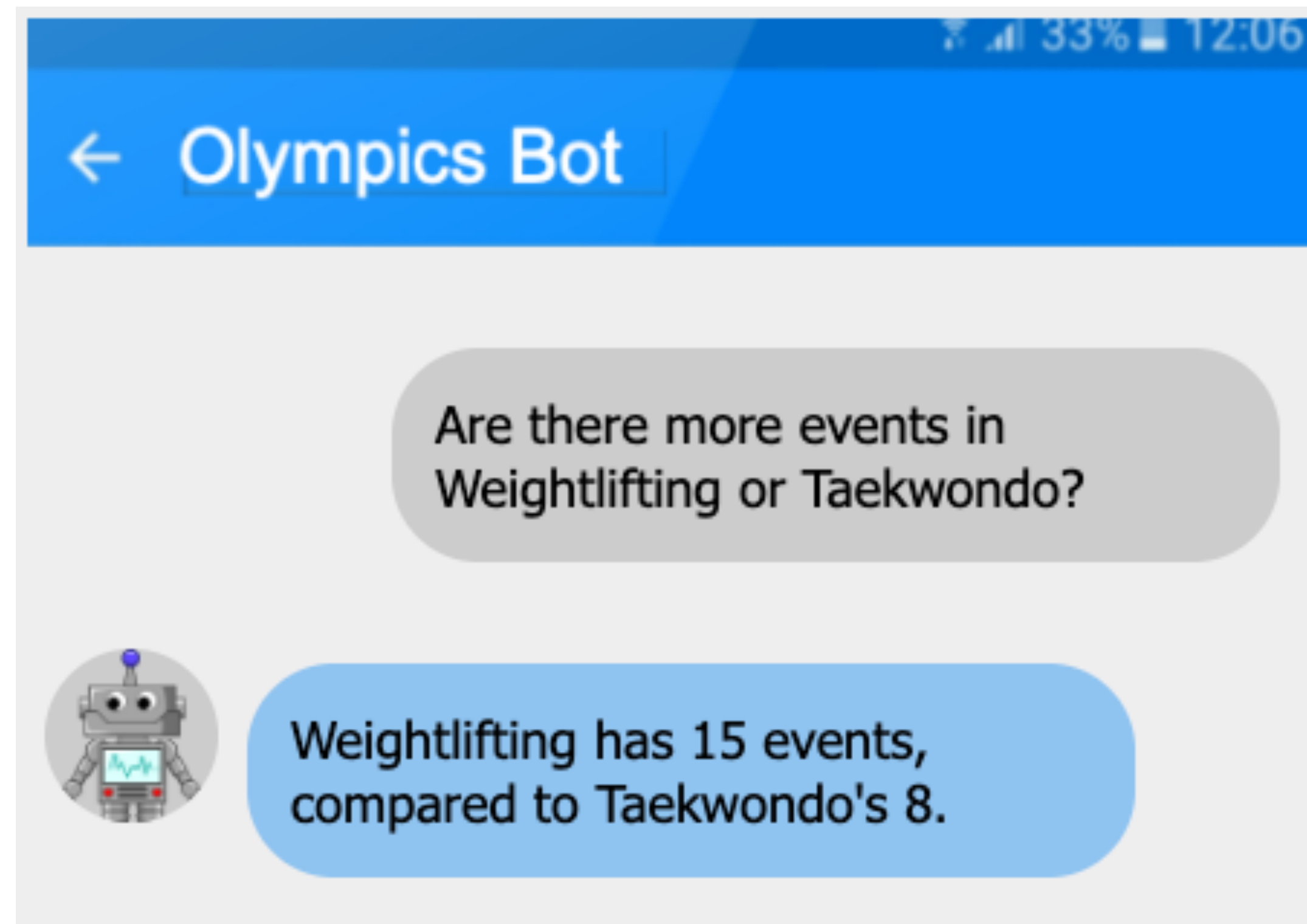


More context preferred over showing only the data points visually.

...and, 41% of participants did not want to see viz in chat

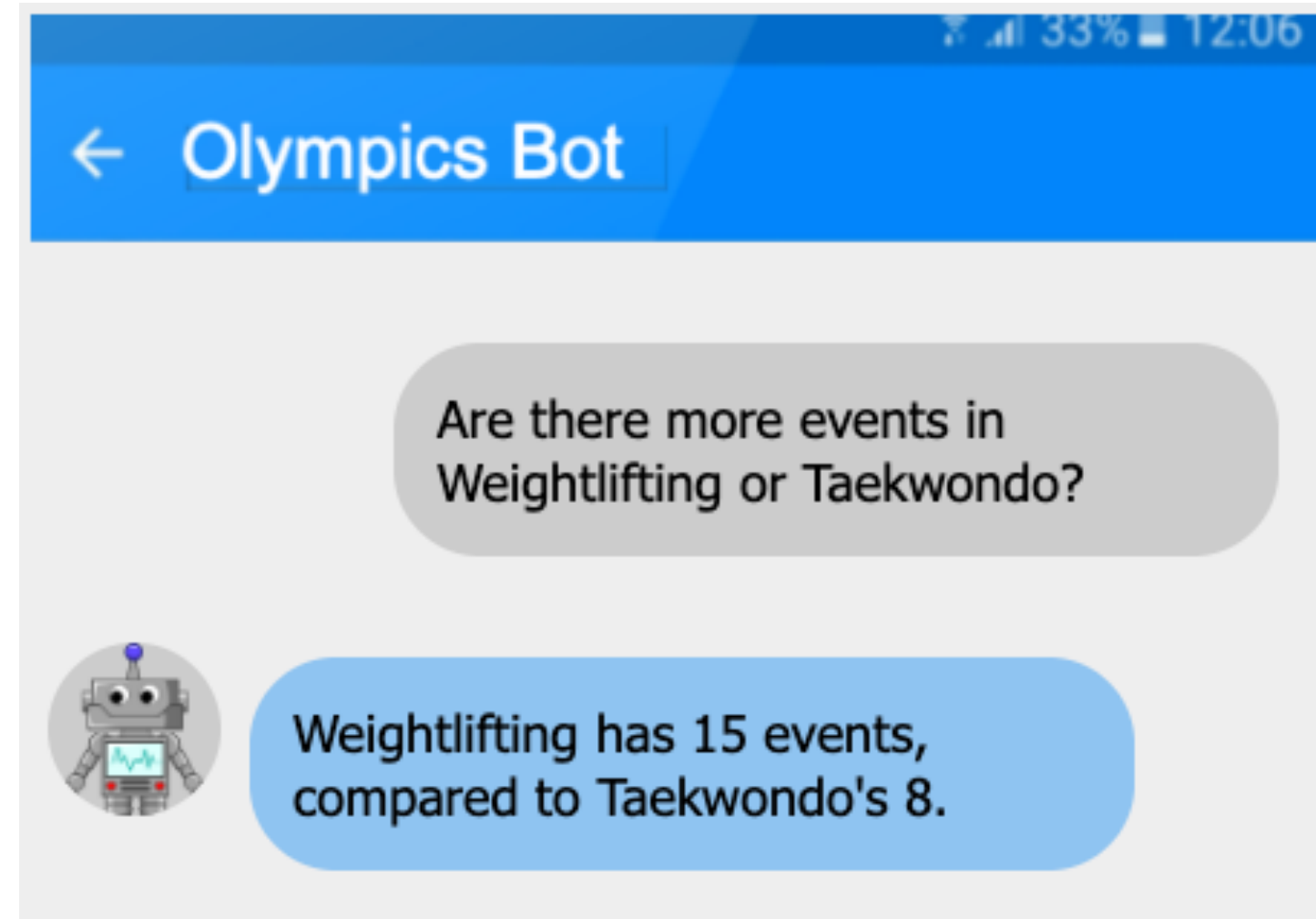


...and, 41% of participants did not want to see viz in chat
They preferred **text alone**



Since then, other studies find similar results

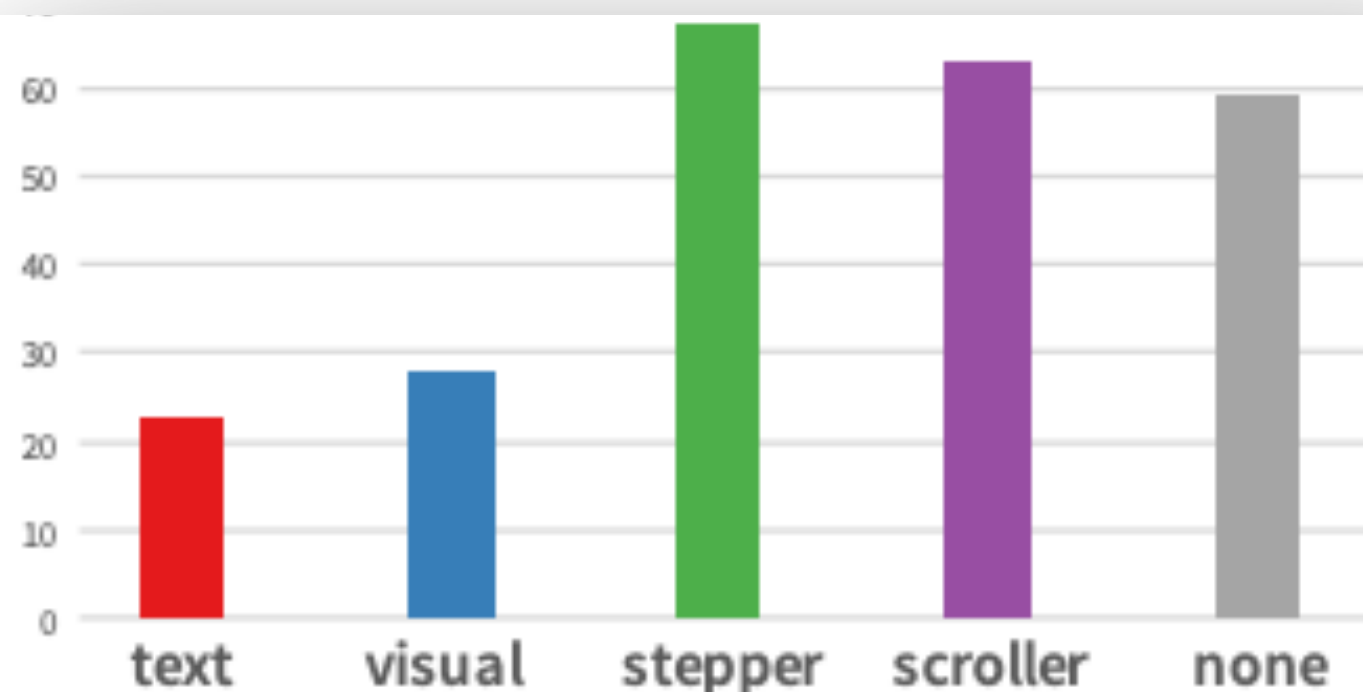
41% did not want to see viz in chat



14% preferred text alone over text + viz

Changes in administration policies have had notable effects on the yearly immigration rate. Starting in 1970, the number of immigrants steadily increased from 50K to around 600K. The Refugee Act in 1981 allowed refugees from several wars to enter the country, which led to a sharp increase in 1981, with 2M arrivals in the following year. Over the next two years, immigration dropped quickly, falling to a little less than 1 million by 1985. By 1990, immigration had increased again, reaching a point of 1.5M. A new administration imposed harsher regulations on immigration, resulting in a steady annual decline, reaching about 100K by 2010.

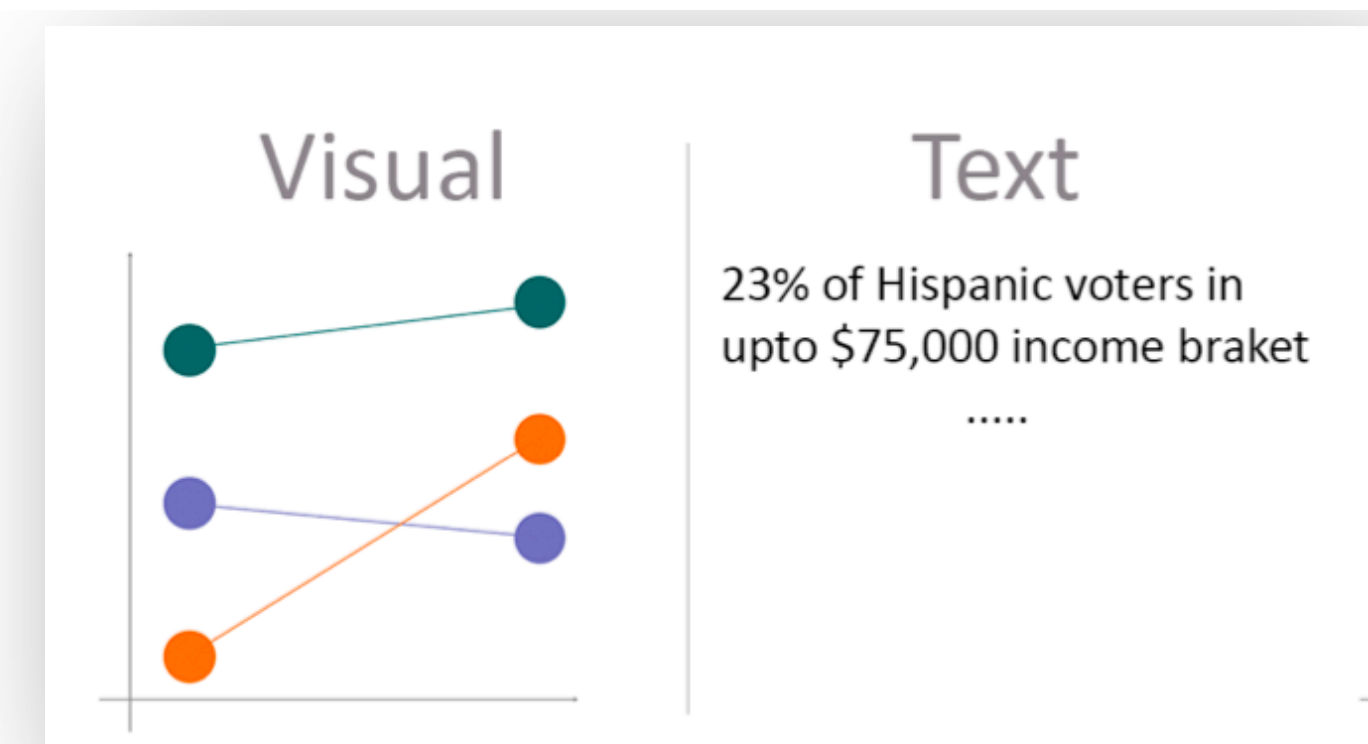
10% prefer text alone



Text alone as accurate as viz; more effective

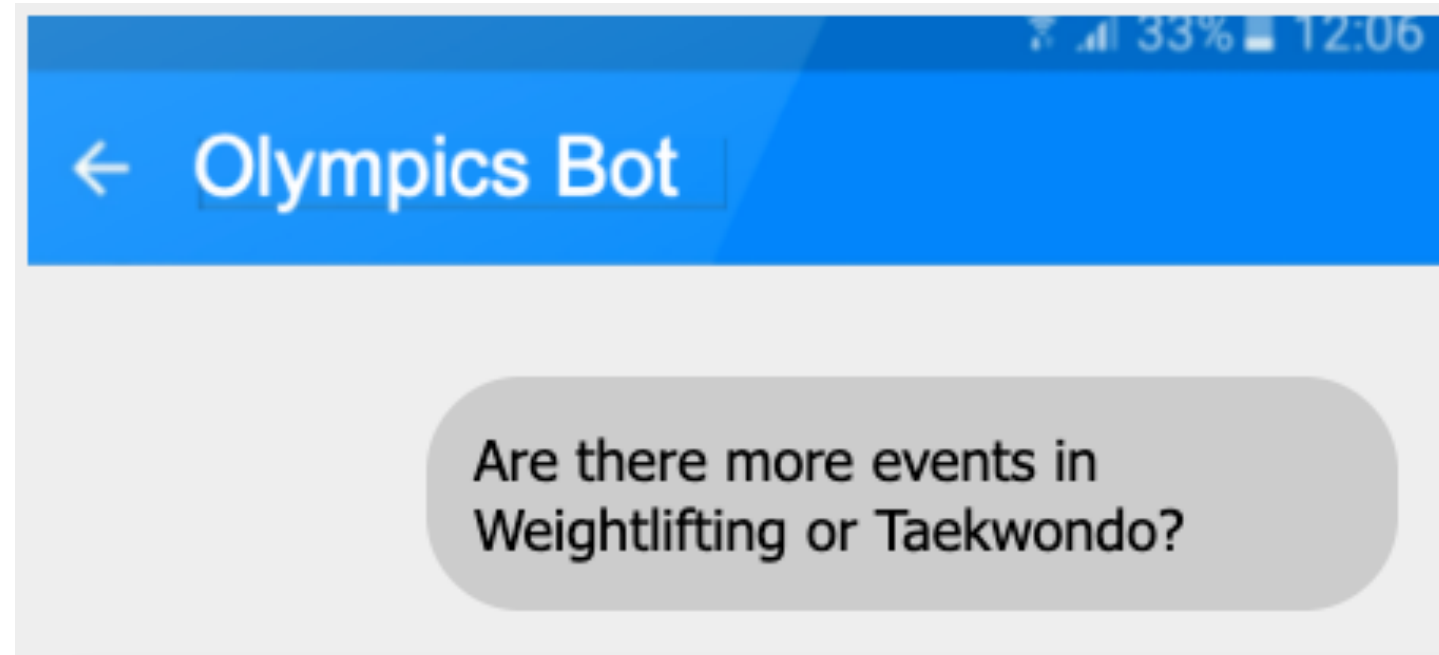
There is a total of 100 college freshmen in the population. Of these freshmen, 30 are enrolled in an introductory entrepreneurship course. Of these 30 freshmen, 20 plan on going into business after graduation and 10 do not. On the other hand, 70 freshmen are not enrolled in an entrepreneurship course. Out of these 70 freshmen, 10 plan on going into business after graduation and 60 do not.

Text alone more accurate for exact recall of stats



Since then, other studies find similar results

41% did not want to see viz in chat

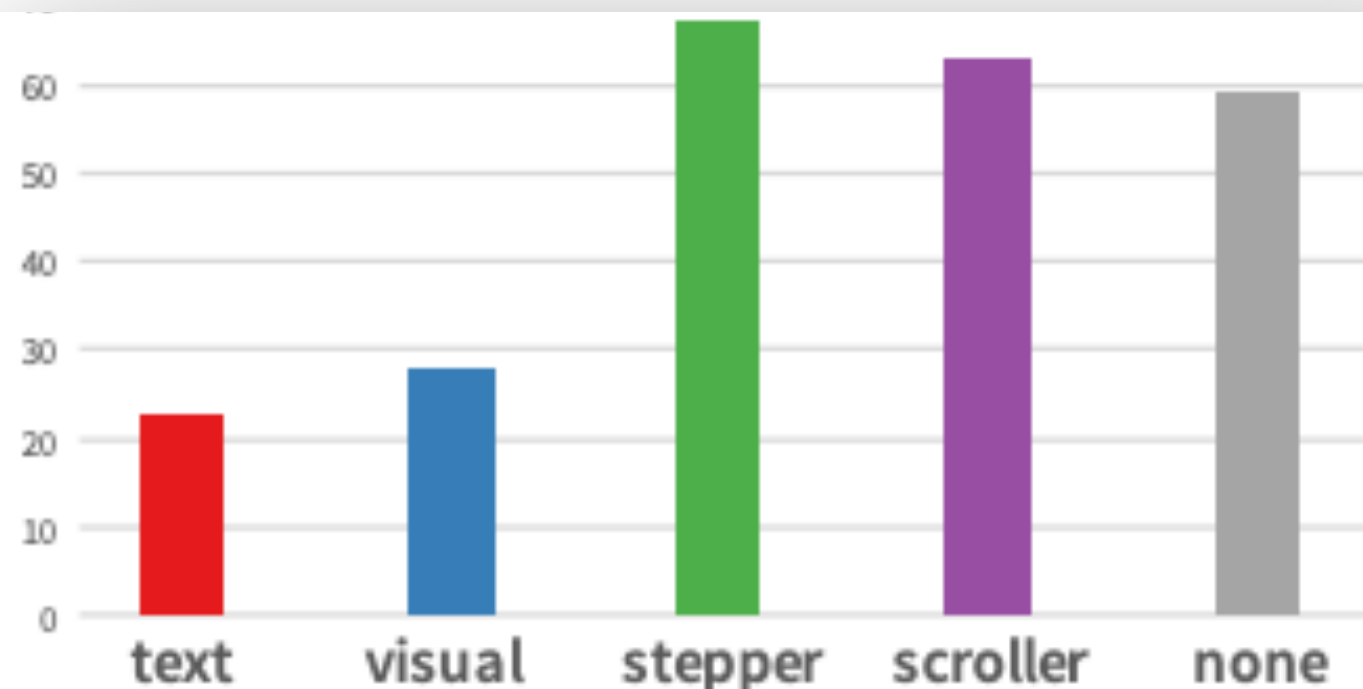


14% preferred text alone over text + viz

Changes in administration policies have had notable effects on the yearly immigration rate. Starting in 1970, the number of immigrants steadily increased from 50K to around 600K. The Refugee Act in 1981 allowed refugees from several wars to enter the country, which led to a sharp increase in 1981, with 2M arrivals in the following year. Over the next two years, immigration

Conversation is **COMPLEX**.

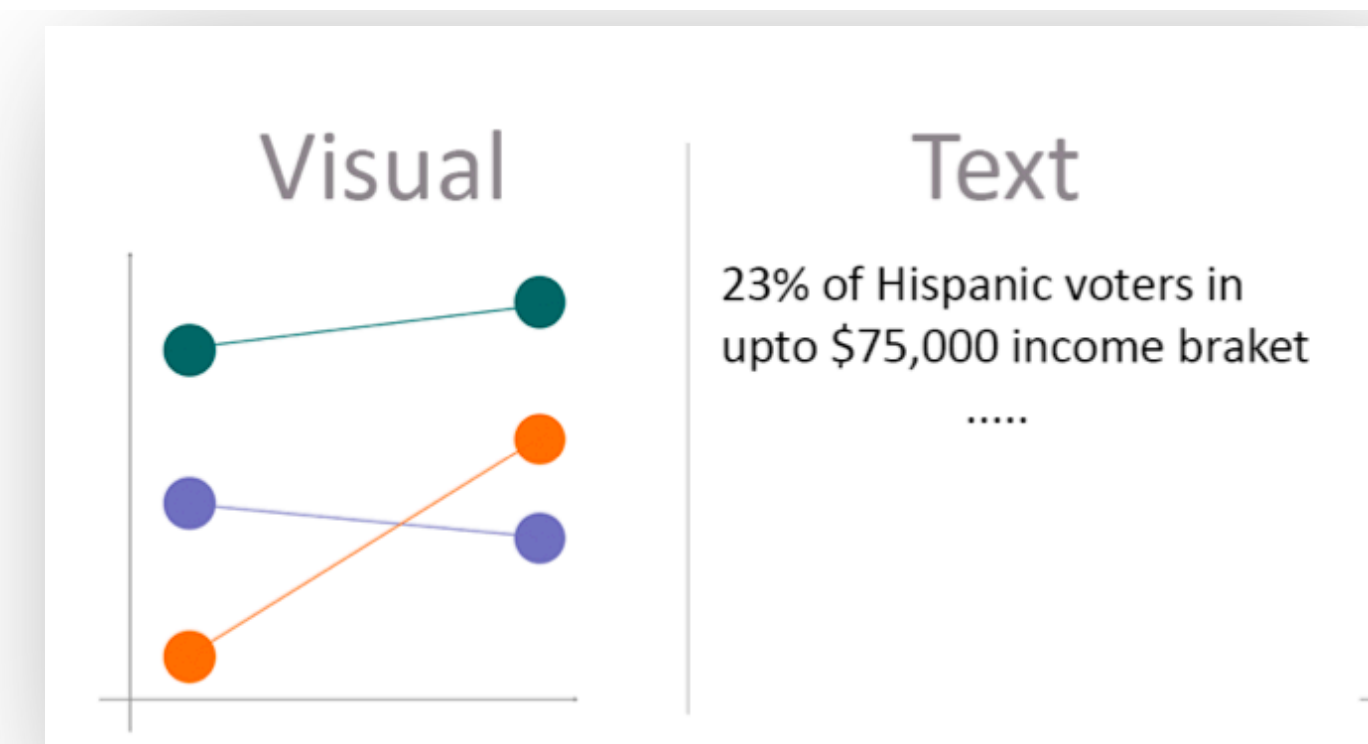
10% prefer text alone



Text alone as accurate as viz; more effective

There is a total of 100 college freshmen in the population. Of these freshmen, 30 are enrolled in an introductory entrepreneurship course. Of these 30 freshmen, 20 plan on going into business after graduation and 10 do not. On the other hand, 70 freshmen are not enrolled in an entrepreneurship course. Out of these 70 freshmen, 10 plan on going into business after graduation and 60 do not.

Text alone more accurate for exact recall of stats



Conversation is *COMPLEX*.

And even moreso when augmenting information artifacts

Designing for Interactive Explanations

Focus: Scientific Articles

Designing for Interactive Explanations

Focus: Scientific Articles

ScholarPhi

PaperPlain

Expandable Abstracts

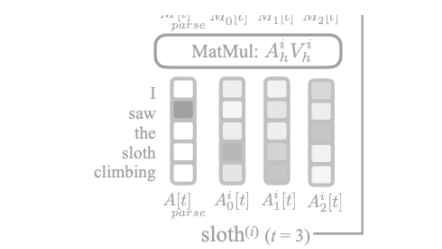


Figure 2: Syntactically-informed self-attention for the query word *sloth*. Attention weights A_{parse} heavily weight the token's syntactic governor, *saw*, in a weighted average over the token values V_{parse} . The other attention heads act as usual, and the attended representations from all heads are concatenated and projected through a feed-forward layer to produce the syntactically-informed representation for *sloth*.

ing hard parameter sharing (§2.3). We simplify optimization and benefit from shared statistical strength derived from highly correlated POS and predicates by treating *main* and *predicate* data as a single task, projecting the joint representation into the joint POS and predicate labels.

Though typical models, which re-encode the sentence for each predicate, can simplify SRL to token-wise tagging, our joint model requires a different approach to classify roles with respect to each predicate. Contextually encoded tokens are projected to distinct *predicate* and *role* embeddings (§2.4), and each predicted predicate is scored with the sequence's role representations us-

trained word embeddings, but we also experiment with supplying pre-trained ELMo representations combined with task-specific learned parameters, which have been shown to substantially improve performance of other SRL models (Peters et al., 2018). For experiments with gold predicates, we concatenate a predicate indicator embedding p_t following previous work (He et al., 2017).

We project² these input embeddings to a representation that is the same size as the output of the self-attention layers. We then add a positional encoding vector computed as a deterministic sinusoidal function of t , since the self-attention has no innate notion of token position.

We feed this token representation as input to a series of J residual multi-head self-attention layers with feed-forward connections. Denoting the

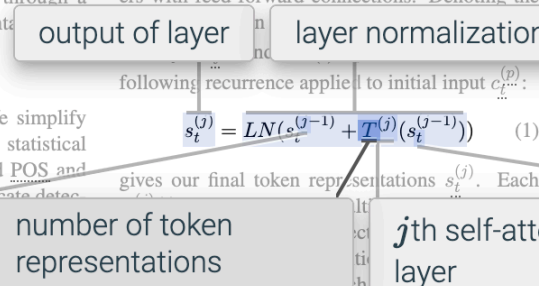
following recurrence applied to initial input $c_t^{(j)}$:

$$s_t^{(j)} = LN(\epsilon_t^{(j-1)} + T^{(j)}(s_t^{(j-1)})) \quad (1)$$

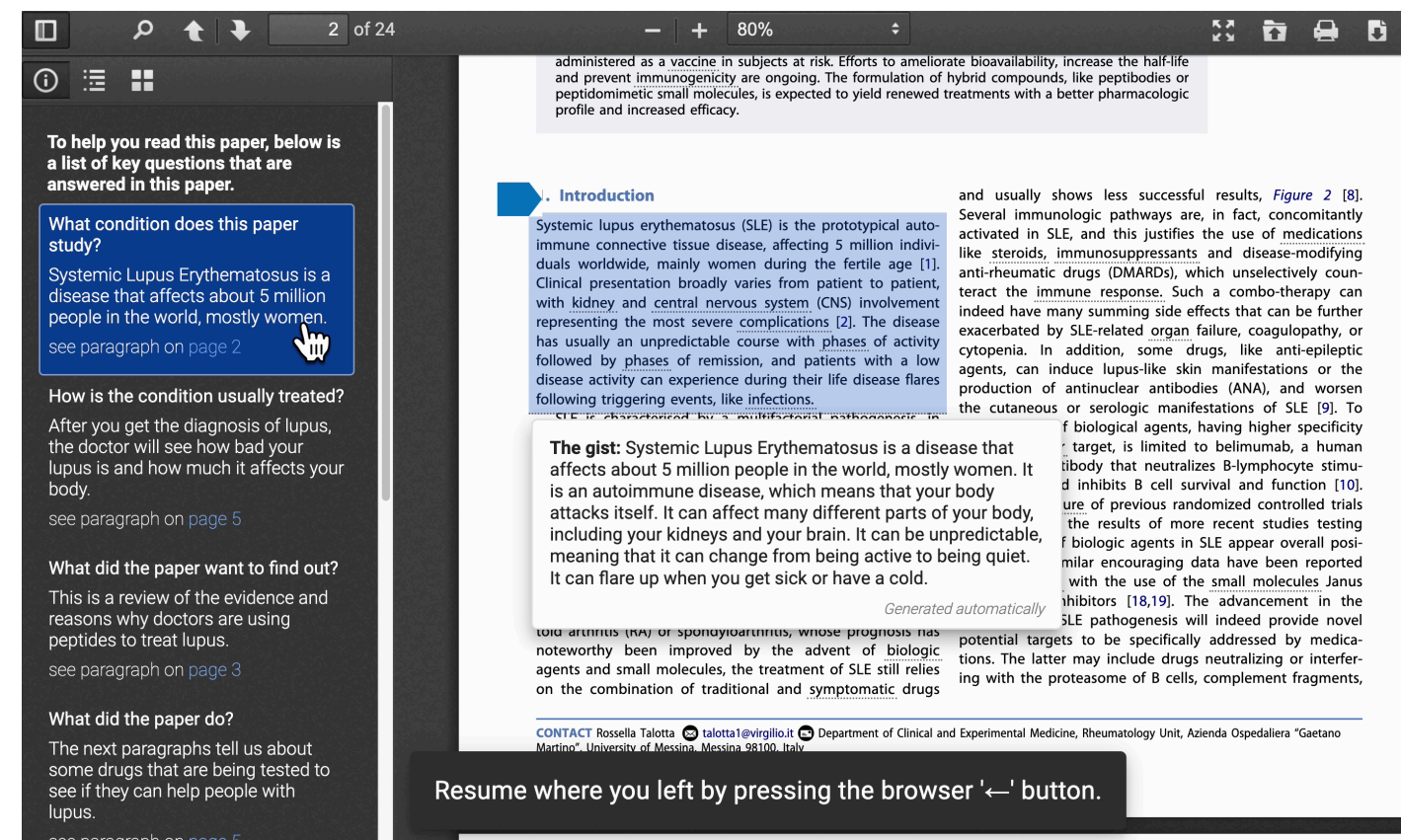
gives our final token representations $s_t^{(J)}$. Each

tion function to attend to all of the tokens in the sequence. This self-attention is performed for each token for each head, and the results of the H self-attentions are concatenated to form the final self-attended representation for each token.

Specifically, consider the matrix $S^{(j-1)}$ of T token representations at layer $j-1$. For each attention head h , we project this matrix into distinct



row of $M_h^{(j)}$



Semi-Oblivious Chase Termination for Linear Existential Rules: An Experimental Study

M. Calautti, Mostafa Milani, Andreas Pieris • arXiv.org • 2023

The chase procedure is a fundamental algorithmic tool in databases that allows us to reason with constraints, such as existential rules, with a plethora of applications. It takes as input a database and a set of constraints, and iteratively completes the database as dictated by the constraints. A key challenge, though, is the fact that it may not terminate, which leads to the problem of checking whether it terminates given a database and a set of constraints. In this work, we focus on the semi-oblivious version of the chase, which is well-suited for practical implementations, and linear existential rules, a central class of constraints with several applications.

Linear existential rules are a type of tuple-generating dependency (TGD) in a database. They are of the form $\forall \vec{x} \forall \vec{y} (\phi(\vec{x}, \vec{y}) \rightarrow \exists \vec{z} \psi(\vec{x}, \vec{z}))$ and have only one quantified variable in the head of the rule. They are called simple-linear if each variable in the rule appears exactly once. Linear existential rules play a role in the chase procedure, which is an algorithm used to produce a universal model of

The phrase "play a role in the chase procedure" refers to the involvement or significance of certain elements or actions in the chase algorithm. In the context of the paper, it refers to the application of tuple-generating dependencies (TGDs) and the addition of atoms to the instance during the chase steps. These actions are essential for the chase procedure to generate a universal model of the database and satisfy all the TGDs.

The chase algorithm is a fundamental algorithmic procedure used in database theory. It takes as input a database and a set of constraints, known as existential rules or tuple-generating dependencies (TGDs), and produces an instance that satisfies all the TGDs. The chase algorithm iteratively applies the TGDs to the database until all the constraints are satisfied.

Requires very careful user interface design and evaluation, combined with NLP and document analysis



Designing for Interactive Explanations

ScholarPhi

SCHOLARPHI

Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols

@ ACM CHI '21.

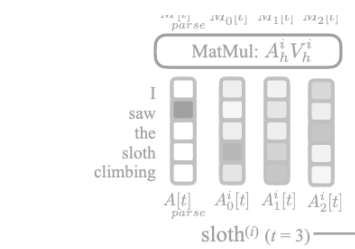


Figure 2: Syntactically-informed self-attention for the query word *sloth*. Attention weights A_{parse} heavily weight the token's syntactic governor, *saw*, in a weighted average over the token values V_{parse} . The other attention heads act as usual, and the attended representations from all heads are concatenated and projected through a feed-forward layer to produce the syntactically-informed representation for *sloth*.

trained word embeddings, but we also experiment with supplying pre-trained ELMo representations combined with task-specific learned parameters, which have been shown to substantially improve performance of other SRL models (Peters et al., 2018). For experiments with gold predicates, we concatenate a predicate indicator embedding p_t following previous work (He et al., 2017).

We project these input embeddings to a representation that is the same size as the output of the self-attention layers. We then add a positional encoding vector computed as a deterministic sinusoidal function of t , since the self-attention has no innate notion of token position.

We feed this token representation as input to a series of J residual multi-head self-attention layers with feed-forward connections. Denoting the

ing hard parameter sharing (§2.3). We simplify optimization and benefit from shared statistical strength derived from highly correlated POS and predicates by treating training and evaluation data as a single task, partitioning the data into training and validation sets. We simplify the joint task into the joint task of predicting POS and predicate labels.

Though typical models, which re-encode the sentence for each predicate, can simplify SRL to token-wise tagging, our joint model requires a different approach to classify roles with respect to each predicate. Contextually encoded tokens are projected to distinct predicate and role embeddings (§2.4), and each predicted predicate is scored with the sequence's role representations using

output of layer

layer normalization

following recurrence applied to initial input $e^{(0)}$:

$$s_t^{(j)} = LN(e^{(j-1)} + \text{FF}(s_t^{(j-1)})) \quad (1)$$

gives our final token representations $s_t^{(J)}$. Each

layer

number of token representations

j th self-attention layer

attention function to attend to all of the tokens in the sequence. This self-attention is performed for each token for each head, and the results of the H self-attentions are concatenated to form the final self-attended representation for each token.

Specifically, consider the matrix $S^{(j-1)}$ of T -token representations at layer $j-1$. For each attention head h , we project this matrix into distinct

row of $M_h^{(j)}$



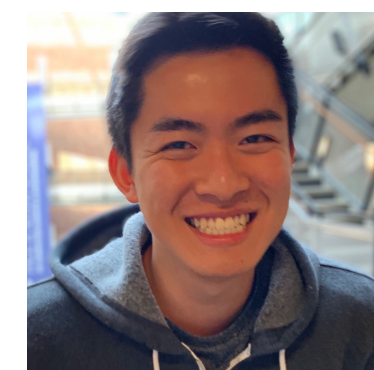
Andrew Head



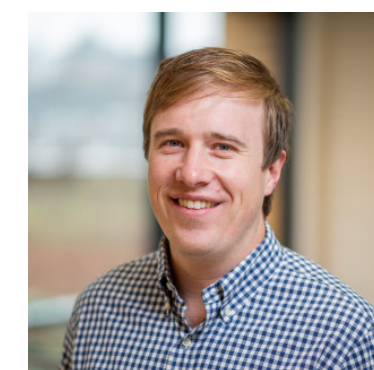
Kyle Lo



Dongyeop Kang



Raymond Fok



Sam Skjonsberg



Daniel S. Weld



Marti A. Hearst



Linguistically-Informed Self-Attention for Semantic Role Labeling

Emma Strubell¹, Patrick Verga¹, Daniel Andor², David Weiss² and Andrew McCallum¹

¹College of Information and Computer Sciences

University of Massachusetts Amherst

{strubell, pat, mccallum}@cs.umass.edu

²Google AI Language

New York, NY

{andor, djweiss}@google.com

Abstract

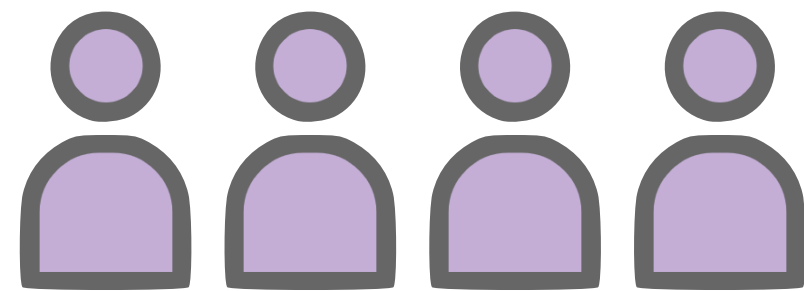
Current state-of-the-art semantic role labeling (SRL) uses a deep neural network with no explicit linguistic features. However, prior work has shown that gold syntax trees can dramatically improve SRL decoding, suggesting the possibility of increased accuracy from explicit modeling of syntax. In this work, we present linguistically-informed self-attention (LISA): a neural network model that combines multi-head self-attention with multi-task learning across dependency parsing, part-of-speech tagging, predicate detection and SRL. Unlike previous models which require significant pre-processing to prepare linguistic features, LISA can incorporate syntax using merely raw tokens as input, encoding the sequence only once to simultaneously perform parsing, predicate detection and role labeling for all predicates. Syntax is incorporated by training one attention head to attend to syntactic parents for each token. Moreover, if a high-quality syntactic parse is already available, it can be beneficially injected at test time without re-training our SRL model.

shown to improve results in challenging downstream tasks such as dialog systems (Tur et al., 2005; Chen et al., 2013), machine reading (Berant et al., 2014; Wang et al., 2015) and translation (Liu and Gildea, 2010; Bazrafshan and Gildea, 2013).

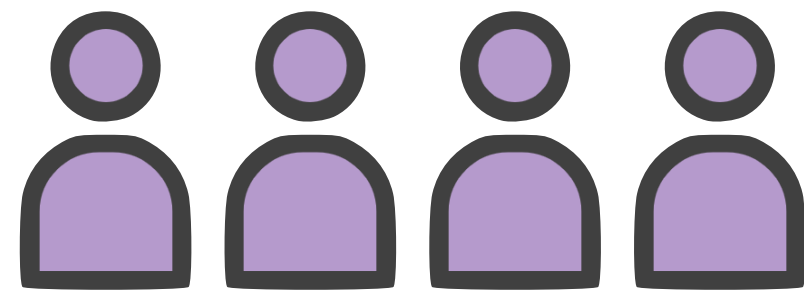
Though syntax was long considered an obvious prerequisite for SRL systems (Levin, 1993; Punyakanok et al., 2008), recently deep neural network architectures have surpassed syntactically-informed models (Zhou and Xu, 2015; Marcheggiani et al., 2017; He et al., 2017; Tan et al., 2018; He et al., 2018), achieving state-of-the-art SRL performance with no explicit modeling of syntax. An additional benefit of these end-to-end models is that they require just raw tokens and (usually) detected predicates as input, whereas richer linguistic features typically require extraction by an auxiliary pipeline of models.

Still, recent work (Roth and Lapata, 2016; He et al., 2017; Marcheggiani and Titov, 2017) indicates that neural network models could see even higher accuracy gains by leveraging syntactic information rather than ignoring it. He et al. (2017)

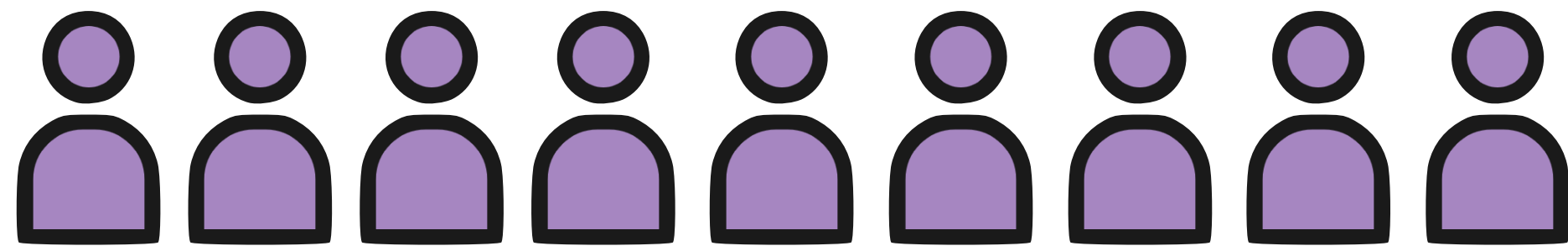
SCHOLARPHI ITERATIVE DESIGN:



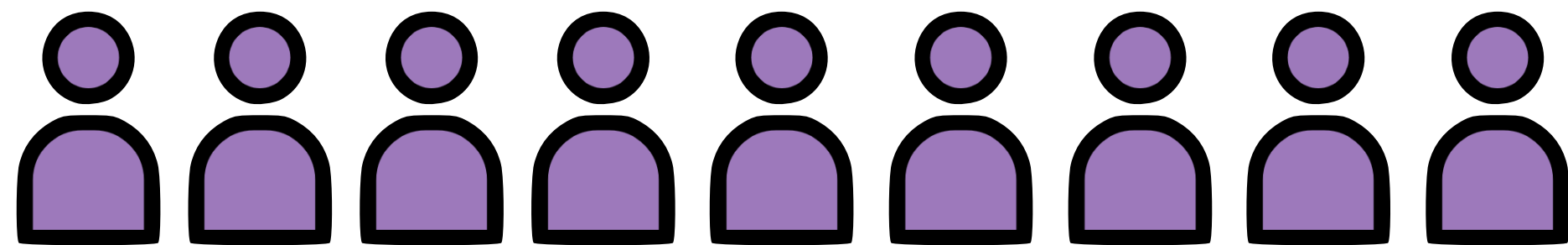
observations



observations



focus group + interviews

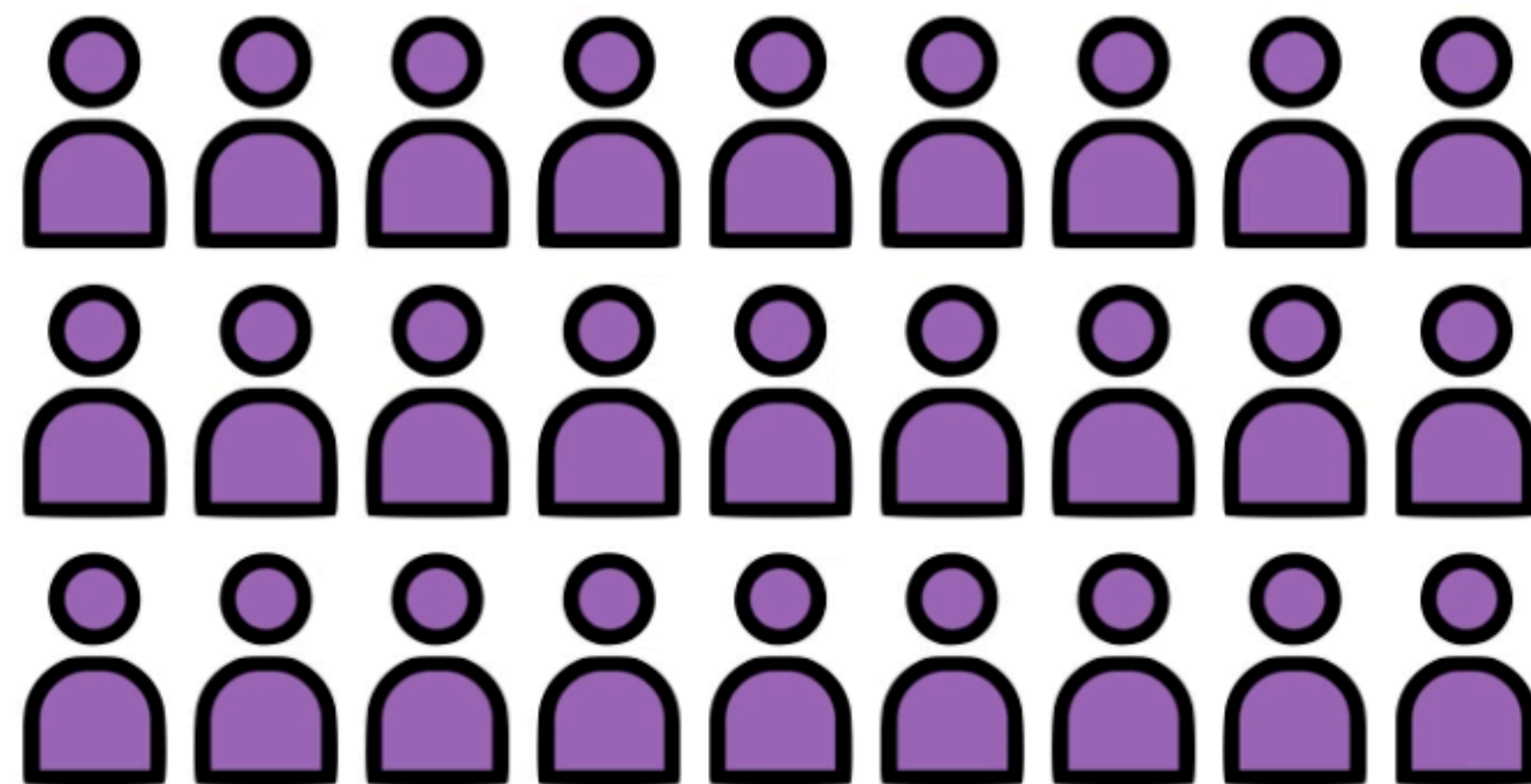


focus group + interviews

4 studies, 24 researchers

Usability study

27 researchers



1 paper



2 tasks

answer questions
about paper
focused reading

Usability findings

When using ScholarPhi (versus a standard PDF reader),
researchers answered questions about a paper...

in **26%** less time,

viewing **53%** less of the paper,

traveling **43%** less distance in the paper.

They found it **significantly easier** to answer the questions, and
were **significantly more confident** in their answers.

Sidebar: Does GPT-4 Solve the Definition Recognition Problem?

It does well, but our specialized algorithm still performs (a bit) better.

	F1
Terms: TADDEX	81.5
Terms: GPT-4	74.8
Definitions: TADDEX	73.6
Definitions: GPT-4	66.1



Anna Martin-Boyle



Dongyeop Kang

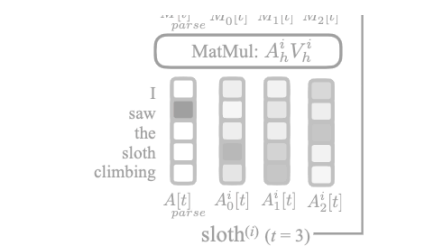


Marti A. Hearst



Designing for Interactive Explanations

ScholarPhi



trained word embeddings, but we also experiment with supplying pre-trained ELMo representations combined with task-specific learned parameters, which have been shown to substantially improve performance of other SRL models (Peters et al., 2018). For experiments with gold predicates, we concatenate a predicate indicator embedding p_t following previous work (He et al., 2017).

We project² these input embeddings to a representation that is the same size as the output of the self-attention layers. We then add a positional encoding vector computed as a deterministic sinusoidal function of t , since the self-attention has no innate notion of token position.

We feed this token representation as input to a series of J residual multi-head self-attention layers with feed-forward connections. Denoting the output of layer j as $s_t^{(j)}$, the following recurrence applies to initial input $s_t^{(0)}$:

output of layer j

layer normalization

$$s_t^{(j)} = LN(s_t^{(j-1)} + T^{(j)}(s_t^{(j-1)})) \quad (1)$$

gives our final token representations $s_t^{(J)}$. Each

layer

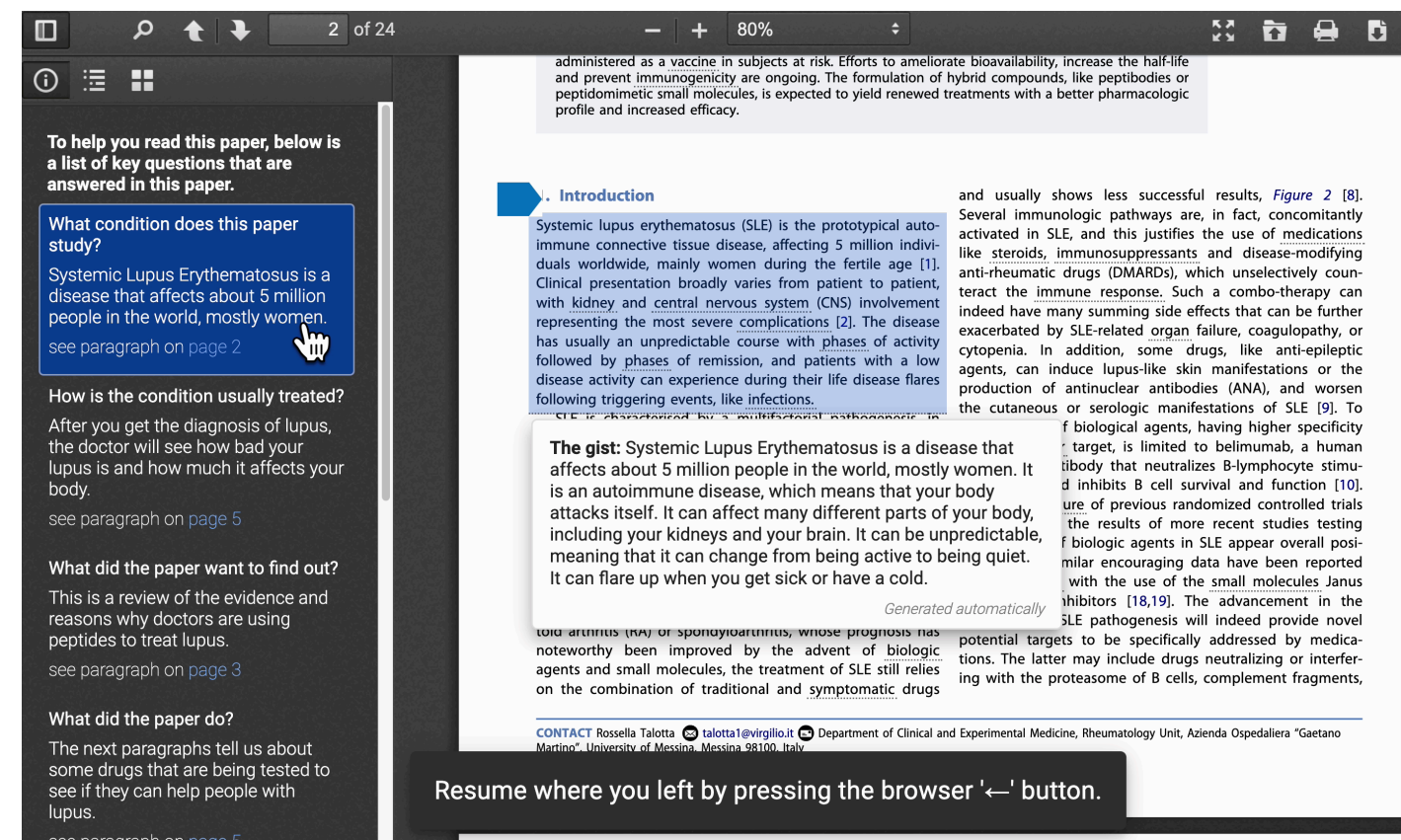
number of token representations

j th self-attention layer

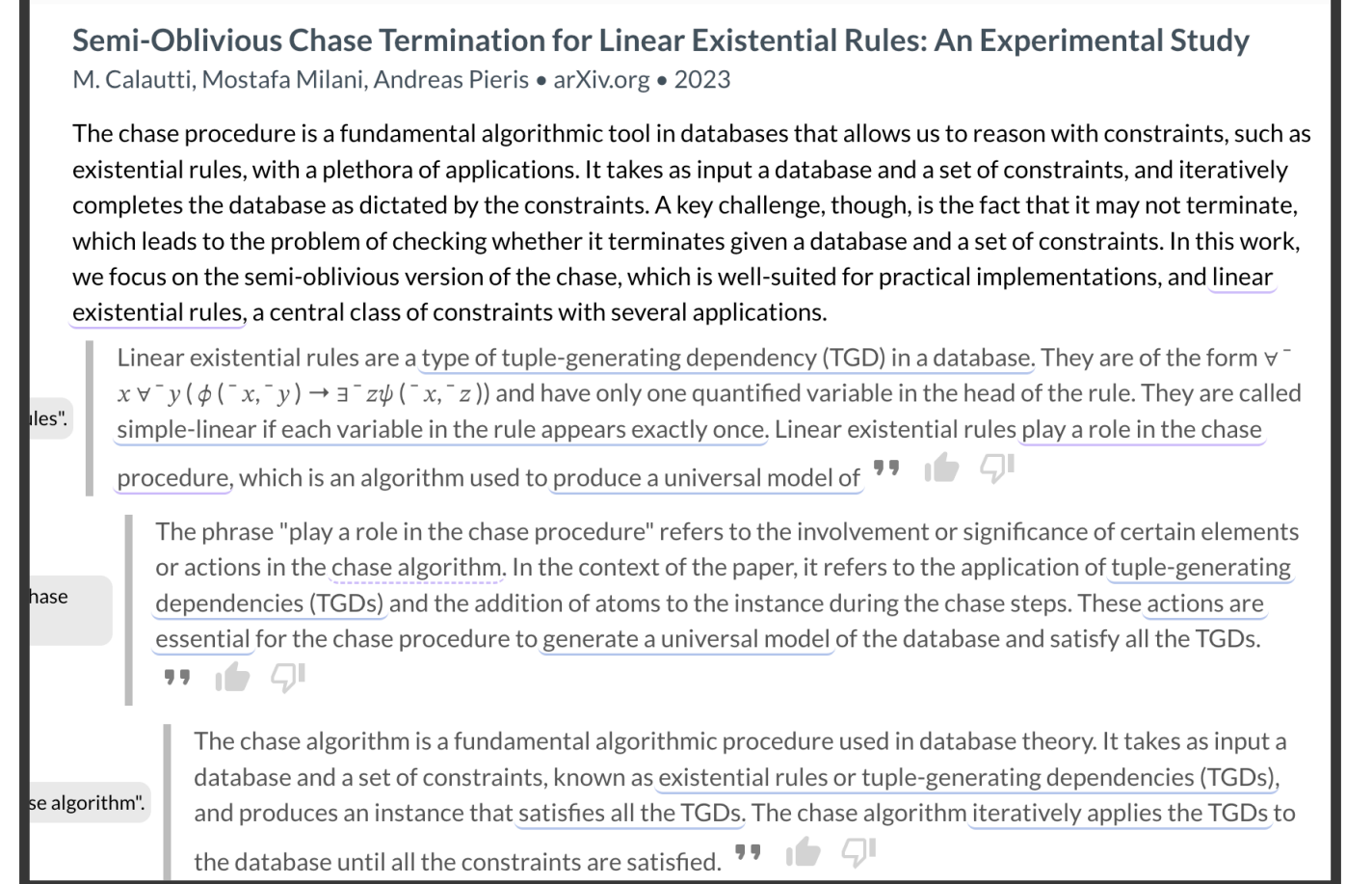
row of $M_h^{(j)}$

Specifically, consider the matrix $S^{(j-1)}$ of T -token representations at layer $j-1$. For each attention head h , we project this matrix into distinct

PaperPlain



Expandable Abstracts

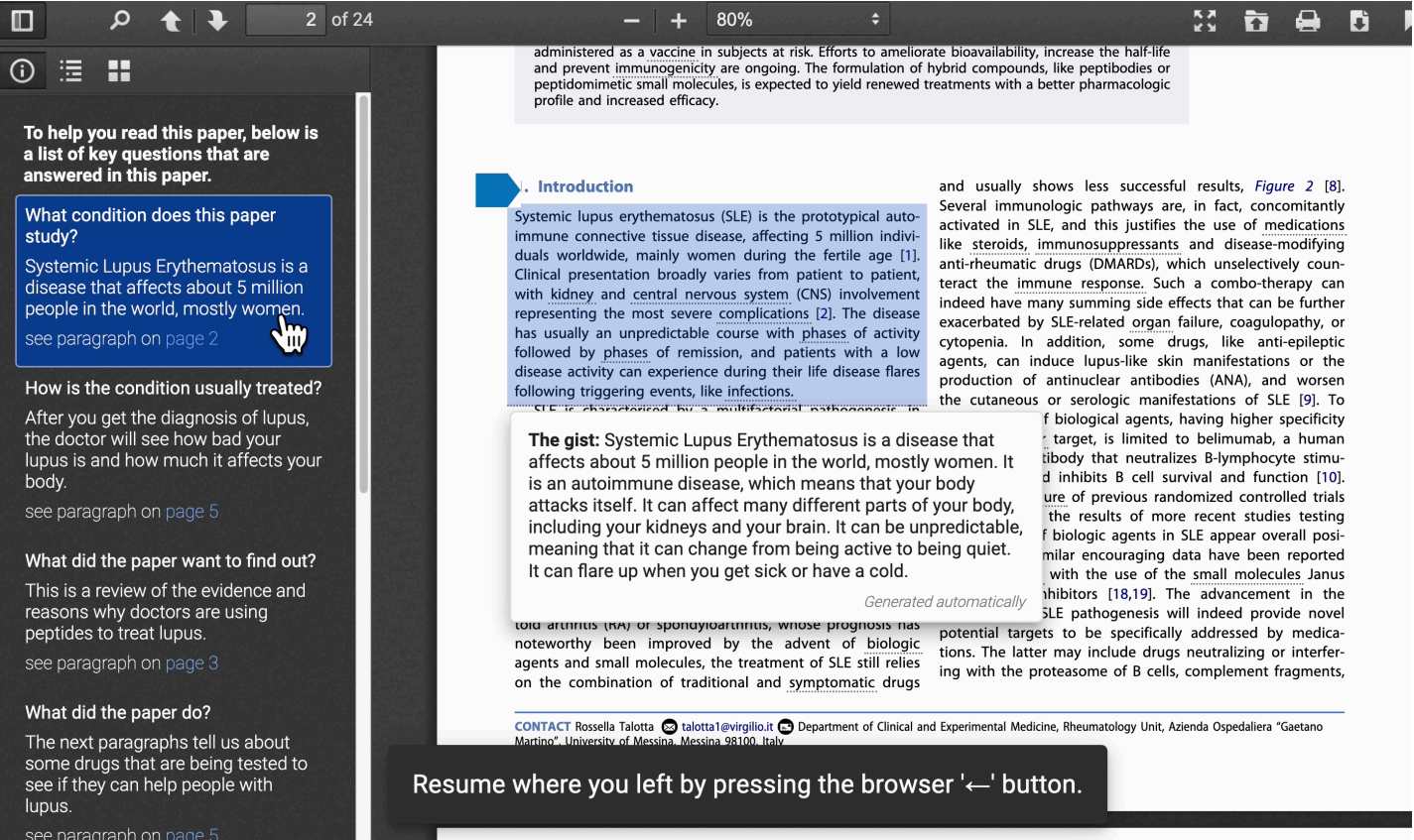


Requires very careful user interface design and evaluation, combined with NLP and document analysis



Designing for Interactive Explanations

PaperPlain



SIMPLIFYING THE COMPLEX

Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing

@ ACM TOCHI '23.



Complex content: how to explain it?

EXPERT OPINION ON INVESTIGATIONAL DRUGS
2020, VOL. 29, NO. 8, 845–867
<https://doi.org/10.1080/13543784.2020.1777983>

REVIEW

Therapeutic peptides for the treatment of systemic lupus erythematosus: a review of the literature

Rossella Talotta^a, Fabiola Atzeni^a and Magdalena Janina Laska^b

^aDepartment of Clinical and Experimental Medicine, Rheumatology Unit, Azienda Ospedaliera, University of Turin, Turin, Italy; ^bDepartment of Rheumatology, Aarhus University Hospital, Aarhus, Denmark

ABSTRACT

Introduction: Studies *in vitro* and *in vivo* have identified several peptides that may be useful in treating systemic lupus erythematosus (SLE). The rationale for their use is based on their production, high potency, target selectivity, low toxicity, and a peculiar mechanism of action, mainly based on the induction of immune tolerance. Three therapeutic peptides have been developed, but they have yielded disappointing results. However, some studies suggest that those with the positivity of anti-dsDNA antibodies, appear more likely to respond.

Areas covered: This review evaluates the potential use of therapeutic peptides in SLE. Our opinion on how they may offer advantages for SLE treatment.

Expert opinion: Given their acceptable safety profile, therapeutic peptides may be considered as traditionally used to treat SLE and this may offer a synergistic and drug-sparing effect.

Add Guiding Questions Determined by User Research

To help you read this paper, below is a list of key questions that are answered in this paper.

What condition does this paper study?

Systemic Lupus Erythematosus is a disease that affects about 5 million people in the world, mostly women.
see paragraph on [page 2](#)



How is the condition usually treated?

After you get the diagnosis of lupus, the doctor will see how bad your lupus is and how much it affects your body.

see paragraph on [page 5](#)

EXPERT OPINION ON INVESTIGATIONAL DRUGS
2020, VOL. 29, NO. 8, 845–867
<https://doi.org/10.1080/13543784.2020.1777983>

REVIEW

Therapeutic peptides for the treatment of systemic therapy

Rossella Talotta^a, Fabiola Atzeni^a and Magdalena Janina Laska^b

^aDepartment of Clinical and Experimental Medicine, Rheumatology Unit, Azienda Ospedale Italy; ^bDepartment of Rheumatology, Aarhus University Hospital, Aarhus, Denmark

ABSTRACT

Introduction: Studies *in vitro* and *in vivo* have identified several peptides that treating systemic lupus erythematosus (SLE). The rationale for their use production, high potency, target selectivity, low toxicity, and a peculiar mechanism mainly based on the induction of immune tolerance. Three therapeutic peptides development, but they have yielded disappointing results. However, some studies those with the positivity of anti-dsDNA antibodies, appear more likely to respond.

Areas covered: This review evaluates the potential use of therapeutic peptides opinion on how they may offer advantages for SLE treatment.

Expert opinion: Given their acceptable safety profile, therapeutic peptides traditionally used to treat SLE and this may offer a synergistic and drug-sparing

Highlight the Answer Passages

To help you read this paper, below is a list of key questions that are answered in this paper.

What condition does this paper study?

Systemic Lupus Erythematosus is a disease that affects about 5 million people in the world, mostly women. see paragraph on [page 2](#)

How is the condition usually treated?

After you get the diagnosis of lupus, the doctor will see how bad your lupus is and how much it affects your body.

see paragraph on [page 5](#)

What did the paper want to find out?

This is a review of the evidence and reasons why doctors are using peptides to treat lupus.

see paragraph on [page 3](#)

What did the paper do?

The next paragraphs tell us about some drugs that are being tested to see if they can help people with lupus.

see paragraph on [page 5](#)

EXPERT OPINION ON INVESTIGATIONAL DRUGS
2020, VOL. 29, NO. 8, 845–867
<https://doi.org/10.1080/13543784.2020.1777983>



REVIEW

Therapeutic peptides for the treatment of systemic lupus erythematosus: a place in therapy

Rossella Talotta^a, Fabiola Atzeni^a and Magdalena Janina Laska^b

^aDepartment of Clinical and Experimental Medicine, Rheumatology Unit, Azienda Ospedaliera "Gaetano Martino", University of Messina, Messina, Italy; ^bDepartment of Rheumatology, Aarhus University Hospital, Aarhus, Denmark

ABSTRACT

Introduction: Studies *in vitro* and *in vivo* have identified several peptides that are potentially useful in treating systemic lupus erythematosus (SLE). The rationale for their use lies in the cost-effective production, high potency, target selectivity, low toxicity, and a peculiar mechanism of action that is mainly based on the induction of immune tolerance. Three therapeutic peptides have entered clinical development, but they have yielded disappointing results. However, some subsets of patients, such as those with the positivity of anti-dsDNA antibodies, appear more likely to respond to these medications.

Areas covered: This review evaluates the potential use of therapeutic peptides for SLE and gives an opinion on how they may offer advantages for SLE treatment.

Expert opinion: Given their acceptable safety profile, therapeutic peptides could be added to agents traditionally used to treat SLE and this may offer a synergistic and drug-sparing effect, especially in selected patient populations. Moreover, they could temporarily be utilized to manage SLE flares, or be administered as a vaccine in subjects at risk. Efforts to ameliorate bioavailability, increase the half-life and prevent immunogenicity are ongoing. The formulation of hybrid compounds, like peptibodies or peptidomimetic small molecules, is expected to yield renewed treatments with a better pharmacologic profile and increased efficacy.

ARTICLE HISTORY

Received 23 January 2020

Accepted 1 June 2020

KEYWORDS

Pathogenesis; systemic lupus erythematosus; therapeutic peptides; treatment; peptibodies; peptides

1. Introduction

Systemic lupus erythematosus (SLE) is the prototypical auto-immune connective tissue disease, affecting 5 million individuals worldwide, mainly women during the fertile age [1]. Clinical presentation broadly varies from patient to patient, with kidney and central nervous system (CNS) involvement representing the most severe complications [2]. The disease

and usually shows less successful results, *Figure 2* [8]. Several immunologic pathways are, in fact, concomitantly activated in SLE, and this justifies the use of medications like steroids, immunosuppressants and disease-modifying anti-rheumatic drugs (DMARDs), which unselectively counteract the immune response. Such a combo-therapy can indeed have many summing side effects that can be further

Provide a Simplified Language Gist

see if they can help people with lupus.

see paragraph on [page 5](#)

What were the new treatment(s), if any, this paper looked into?

A peptide is a small molecule made up of amino acids, which are the building blocks of proteins.

see paragraph on [page 3](#)

What did the paper find?

In the future, scientists will make new drugs that will help people with lupus.

see paragraph on [page 18](#), [page 20](#), [page 19](#)

Are the findings different depending on a person's demographics?

These new drugs are designed to target the bad proteins that cause SLE.

see paragraph on [page 4](#), [page 17](#)

What are the limitations of the findings?

Even though the medicine worked well in the lab, it did not work as well in real life

see paragraph on [page 18](#)

cytic cells may unleash, through an impaired phagocytosis, the cascade of events leading to type I IFN production, and amplify, through this pathway, inflammation and tissue damage.

Conclusion

Better knowledge of the pathogenesis of SLE is expected to enrich the therapeutic armamentarium and facilitate the management of the disease. The use of peptides, specifically designed to target SLE-related epitopes or crucial pathways, may represent a novel fascinating opportunity. Given their good safety profile and immunomodulatory properties, therapeutic peptides could be added to standard of care, and, perhaps, allow the sparing of conventional drugs. In addition, their prescription might be tailored to specific subsets of patients having the highest likelihood of response. Nevertheless, despite the successful results observed in pre-clinical studies, RCTs showed a controversial efficacy profile concerning the use of these compounds in SLE. It is expected that future research, aiming at the amelioration of their physicochemical properties and at the improvement in the design of clinical trials, will bring more encouraging data on this innovative therapeutic panorama.

Expert opinion

The treatment of SLE still relies on a conservative approach, combining multiple unselective immunosuppressive agents [8] and, consecutively, increasing the risk of unwanted side effects. Unlike other rheumatic diseases, the licensed use of biologic agents, which electively inhibit a specific target, has been solely limited to belimumab. Rituximab failed to achieve the primary endpoints in RCTs conducted in SLE patients [12,13,143] but, due to encouraging real-life data [11], its off-label use is advised in resistant severe manifestations [8]. The potential use of novel biological agents and small molecules in SLE is still an object of clinical investigation [10,18]. Furthermore, several preliminary data on other small

sequence by means of recombinant DNA technology and then modified in order to enhance their pharmacologic properties. Most of therapeutic peptides designed for SLE treatment are synthesized on the basis of immunodominant epitope sequences that are pathogenic in SLE alone. Therefore, they exert an immunomodulatory effect on auto-reactive pDC and lymphocytes, without affecting, instead, the immune response against pathogens [26].

To date, no therapeutic peptide has been licensed and marketed for the use in SLE patients. The 21-mer peptide P140 is the only one entering phase III RCTs, and, despite controversial results [36,40,41,43], its development is still ongoing. Edratide, synthesized on the basis of the hCDR1 expressing the major idiotype 16/6 Id, showed promising results in preclinical studies [48,50,54,55], but failed to meet the primary endpoint in a phase II RCT [57], with the following interruption of further clinical development. Both the two peptides appeared more effective in anti-dsDNA seropositive patients, and, although data are lacking, a beneficial role may be supposed in those SLE manifestations related to anti-dsDNA antibodies, such as glomerulonephritis.

The gist: A drug that targets a specific part of the immune system is being tested to see if it can help people with a disease called lupus. The drug is not yet approved, but it has been tested in people with lupus and seems to help them.

Generated automatically

may be translated in humans is an uncertain issue that needs to be addressed in future research.

The discovery of novel molecular targets is expected to enrich, in the next years, the panorama of therapeutic peptides for SLE, among which CXCR4 and STING antagonists and virus-derived peptides seem promising candidates [116,119,135].

Meanwhile, research is focusing on the optimization of the physicochemical structure of preexisting peptides with the intention to improve their pharmacologic properties, including bioavailability and half-life. One of these efforts consists of the

Resume where you left by pressing the browser's '←' button.

Usability Study Results

Baseline PDF reader: Most participants read papers linearly and spent substantial time in dense sections with limited important information.

Paper Plain: All participants reached the end of the paper.

Section gists and term definitions helped read dense passages.

The questions & answer gists supported quick navigation and helped decide which sections to read

Designing for Interactive Explanations

ScholarPhi

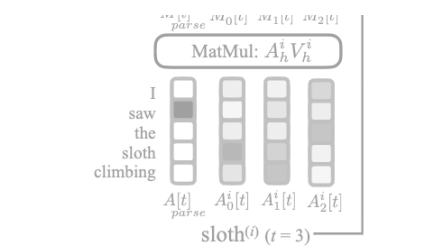


Figure 2: Syntactically-informed self-attention for the query word *sloth*. Attention weights A_{parse} heavily weight the token's syntactic governor, *saw*, in a weighted average over the token values V_{parse} . The other attention heads act as usual, and the attended representations from all heads are concatenated and projected through a feed-forward layer to produce the syntactically-informed representation for *sloth*.

ing hard parameter sharing (§2.3). We simplify optimization and benefit from shared statistical strength derived from highly correlated POS and predicates by treating *main* and *predicate* data as a single task, projecting the joint representation into the joint POS and predicate labels.

Though typical models, which re-encode the sentence for each predicate, can simplify SRL to token-wise tagging, our joint model requires a different approach to classify roles with respect to each predicate. Contextually encoded tokens are projected to distinct *predicate* and *role* embeddings (§2.4), and each predicted predicate is scored with the sequence's role representations us-

trained word embeddings, but we also experiment with supplying pre-trained ELMo representations combined with task-specific learned parameters, which have been shown to substantially improve performance of other SRL models (Peters et al., 2018). For experiments with gold predicates, we concatenate a predicate indicator embedding p_t following previous work (He et al., 2017).

We project² these input embeddings to a representation that is the same size as the output of the self-attention layers. We then add a positional encoding vector computed as a deterministic sinusoidal function of t , since the self-attention has no innate notion of token position.

We feed this token representation as input to a series of J residual multi-head self-attention layers with feed-forward connections. Denoting the

following recurrence applied to initial input $c_t^{(0)}$:

$$s_t^{(j)} = LN(\epsilon_t^{(j-1)} + T^{(j)}(s_t^{(j-1)})) \quad (1)$$

gives our final token representations $s_t^{(J)}$. Each

attention function to attend to all of the tokens in the sequence. This self-attention is performed for each token for each head, and the results of the H self-attentions are concatenated to form the final self-attended representation for each token.

Specifically, consider the matrix $S^{(j-1)}$ of T token representations at layer $j-1$. For each attention head h , we project this matrix into distinct

output of layer

layer normalization

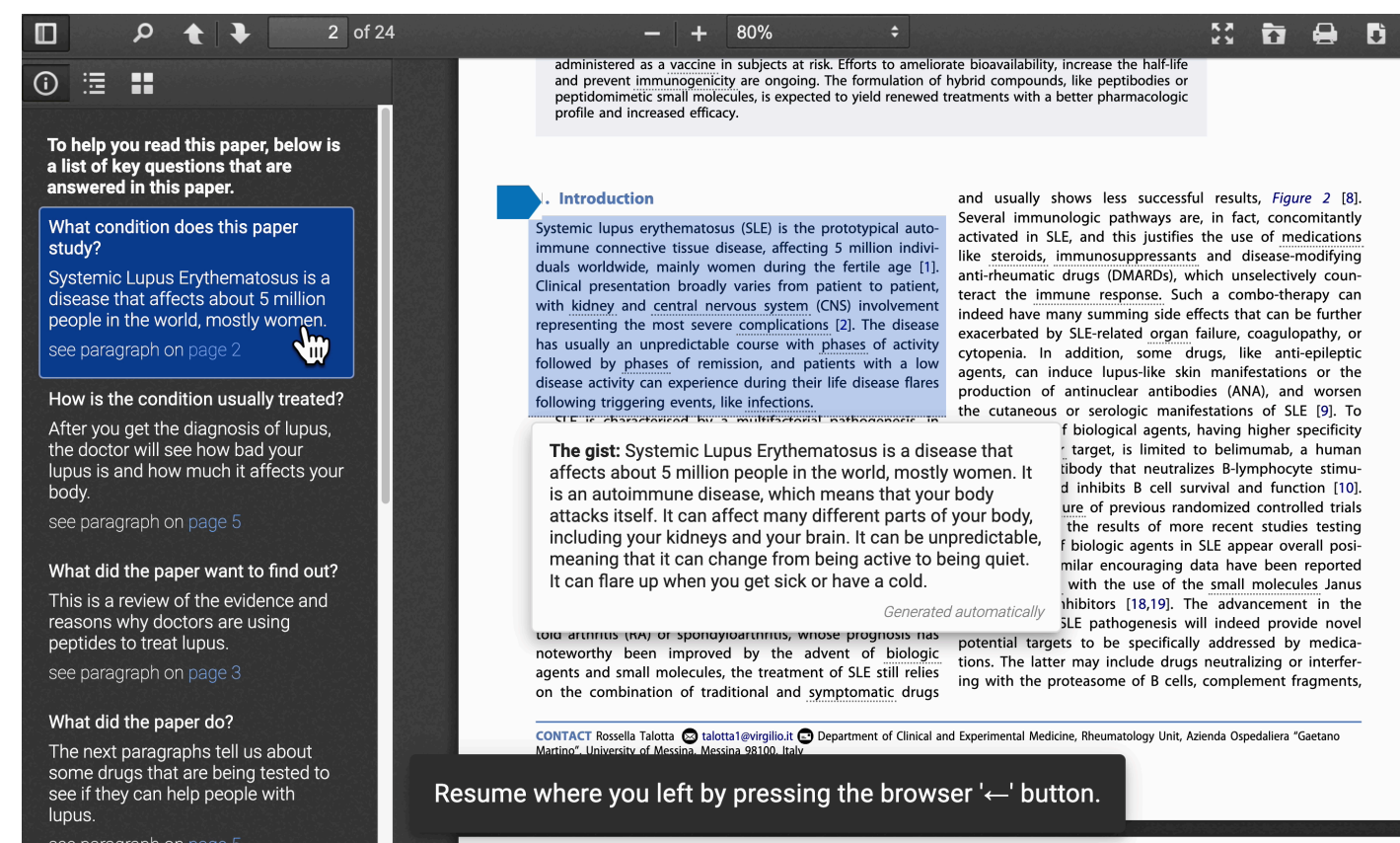
layer

number of token representations

jth self-attention layer

row of $M_h^{(j)}$

PaperPlain



Expandable Abstracts

Semi-Oblivious Chase Termination for Linear Existential Rules: An Experimental Study

M. Calautti, Mostafa Milani, Andreas Pieris • arXiv.org • 2023

The chase procedure is a fundamental algorithmic tool in databases that allows us to reason with constraints, such as existential rules, with a plethora of applications. It takes as input a database and a set of constraints, and iteratively completes the database as dictated by the constraints. A key challenge, though, is the fact that it may not terminate, which leads to the problem of checking whether it terminates given a database and a set of constraints. In this work, we focus on the semi-oblivious version of the chase, which is well-suited for practical implementations, and linear existential rules, a central class of constraints with several applications.

Linear existential rules are a type of tuple-generating dependency (TGD) in a database. They are of the form $\forall \vec{x} \forall \vec{y} (\phi(\vec{x}, \vec{y}) \rightarrow \exists \vec{z} \psi(\vec{x}, \vec{z}))$ and have only one quantified variable in the head of the rule. They are called simple-linear if each variable in the rule appears exactly once. Linear existential rules play a role in the chase procedure, which is an algorithm used to produce a universal model of

The phrase "play a role in the chase procedure" refers to the involvement or significance of certain elements or actions in the chase algorithm. In the context of the paper, it refers to the application of tuple-generating dependencies (TGDs) and the addition of atoms to the instance during the chase steps. These actions are essential for the chase procedure to generate a universal model of the database and satisfy all the TGDs.

The chase algorithm is a fundamental algorithmic procedure used in database theory. It takes as input a database and a set of constraints, known as existential rules or tuple-generating dependencies (TGDs), and produces an instance that satisfies all the TGDs. The chase algorithm iteratively applies the TGDs to the database until all the constraints are satisfied.

Requires very careful user interface design and evaluation, combined with NLP and document analysis



Designing for Interactive Explanations

Expandable Abstracts

Semi-Oblivious Chase Termination for Linear Existential Rules: An Experimental Study

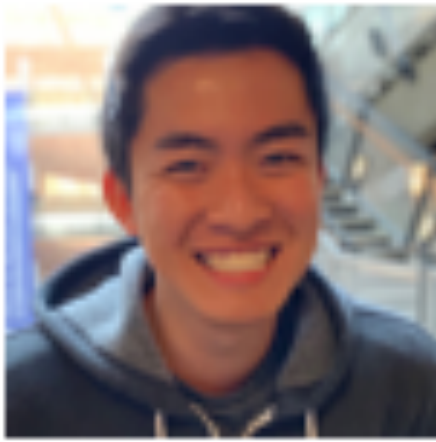
M. Calautti, Mostafa Milani, Andreas Pieris • arXiv.org • 2023

The chase procedure is a fundamental algorithmic tool in databases that allows us to reason with constraints, such as existential rules, with a plethora of applications. It takes as input a database and a set of constraints, and iteratively completes the database as dictated by the constraints. A key challenge, though, is the fact that it may not terminate, which leads to the problem of checking whether it terminates given a database and a set of constraints. In this work, we focus on the semi-oblivious version of the chase, which is well-suited for practical implementations, and linear existential rules, a central class of constraints with several applications.

Linear existential rules are a type of tuple-generating dependency (TGD) in a database. They are of the form $\forall \vec{x} \forall \vec{y} (\phi(\vec{x}, \vec{y}) \rightarrow \exists \vec{z} \psi(\vec{x}, \vec{z}))$ and have only one quantified variable in the head of the rule. They are called simple-linear if each variable in the rule appears exactly once. Linear existential rules play a role in the chase procedure, which is an algorithm used to produce a universal model of ” ” ”

The phrase "play a role in the chase procedure" refers to the involvement or significance of certain elements or actions in the chase algorithm. In the context of the paper, it refers to the application of tuple-generating dependencies (TGDs) and the addition of atoms to the instance during the chase steps. These actions are essential for the chase procedure to generate a universal model of the database and satisfy all the TGDs. ” ” ”

The chase algorithm is a fundamental algorithmic procedure used in database theory. It takes as input a database and a set of constraints, known as existential rules or tuple-generating dependencies (TGDs), and produces an instance that satisfies all the TGDs. The chase algorithm iteratively applies the TGDs to the database until all the constraints are satisfied. ” ” ”



Raymond Fok



Daniel S. Weld



Proceedings of VLDB 2023

Filter papers...

Semantics-aware Dataset Discovery from Data Lakes with Contextualized Column-based Representation Learning

Grace Fan, Jin Wang, Yuliang Li, Dan Zhang, Renée J. Miller

Semi-Oblivious Chase Termination for Linear Existential Rules: An Experimental Study

M. Calautti, Mostafa Milani, Andreas Pieris

Serving and Optimizing Machine Learning Workflows on Heterogeneous Infrastructures

Yongji Wu, Matthew Lentz, Danyang Zhuo, Yao Lu

SkinnerMT: Parallelizing for Efficiency and Robustness in Adaptive Query Processing on Multicore Platforms

Ziyun Wei, Immanuel Trummer

Space-Efficient Random Walks on Streaming Graphs

Serafeim Papadias, Zoi Kaoudi, Jorge-Arnulfo Quiané-Ruiz, V. Markl

Spade: A Real-Time Fraud Detection Framework on Evolving Graphs (Complete Version)

Jiaxin Jiang, Yuan Li, Bin He, Bryan Hooi, J. Chen, Johan Kok Zhi Kang

Sparkly: A Simple yet Surprisingly Strong TF/IDF Blocker for Entity Matching

Derek Paulsen, Yash Govind, A. Doan

SyncSignature: A Simple, Efficient, Parallelizable Framework for Tree Similarity Joins

Nikolai Karpov, Qin Zhang

Talk Outline

01

Generative AI / 'Natural' UIs

02

Designing for Conversations

03

Designing for Automation Bias

03

Designing for Automation Bias

Problem: Paying Attention with Automation (Automation Bias)



Reading medical scans, piloting airplanes, credit scoring, self-driving cars ...

Problem: Automation Bias

Automation Bias: “The tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing.” (Mosier & Skitka 1996, Lyell & Coeira 2017)

System accuracy: System accuracy is known to influence both user expectations of a system as well as user interaction with the system (Chancey et al., 2016).

Trust Calibration: Trust is a nonlinear function of automation performance and the dynamic interaction between the operator and the automation (Lee & See, 2004)



Causes of Automation Bias

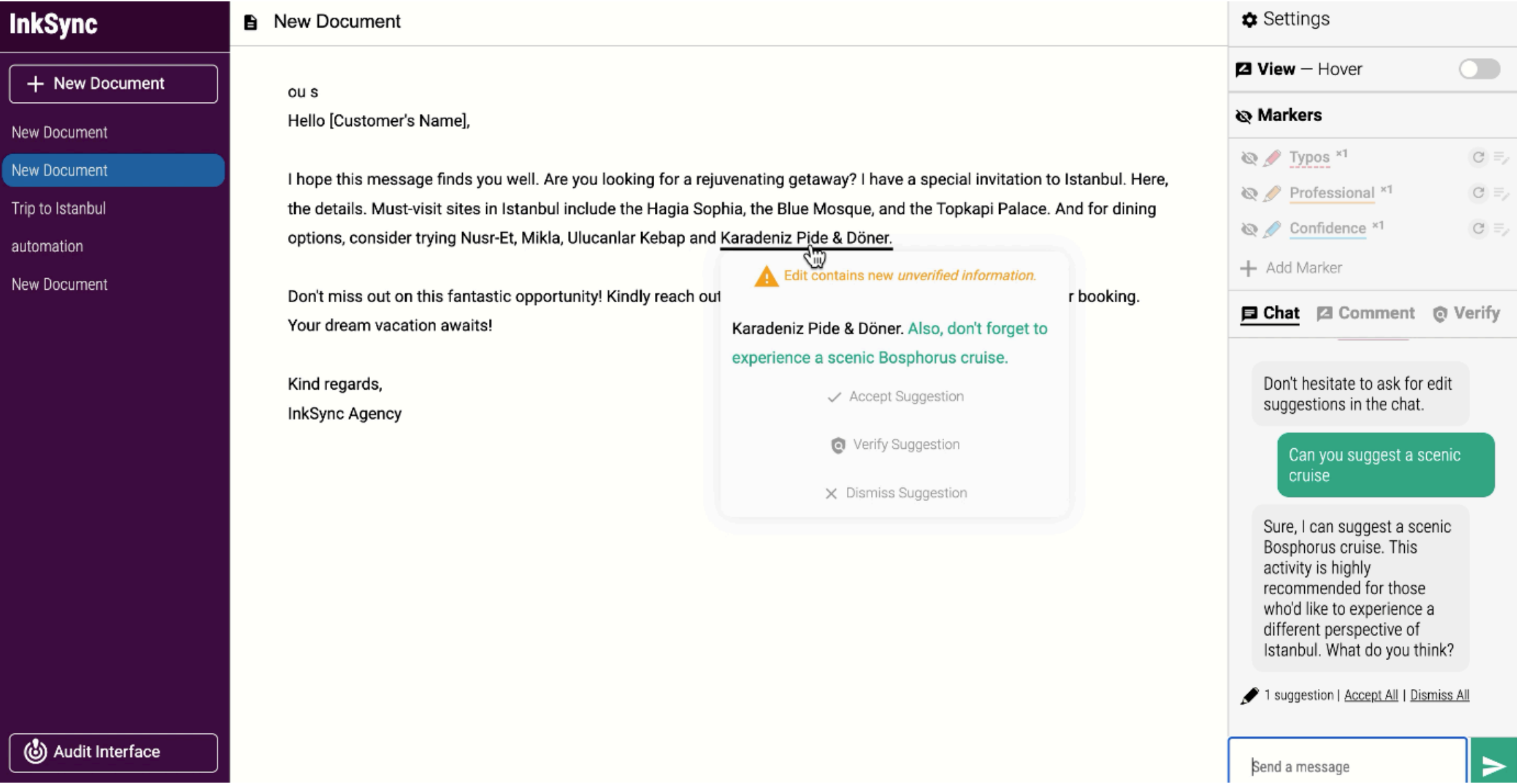
Parasuraman and Manzey [21] identify causes of automation bias (from Gsenger & Strle 2021), including:

People prefer to reduce their cognitive load and thus decide according to simple decision rules and comprehensive heuristics

Automated systems are perceived as powerful agents, believed to have more analytic capabilities than people, and thus they are trusted more.

Designing Conversations for Interactive Writing

InkSync



Beyond the Chat: Editable and Verifiable Text Editing with LLMs

(under review)



Philippe Laban



Jessie Vig



Marti Hearst



Caiming Xiong

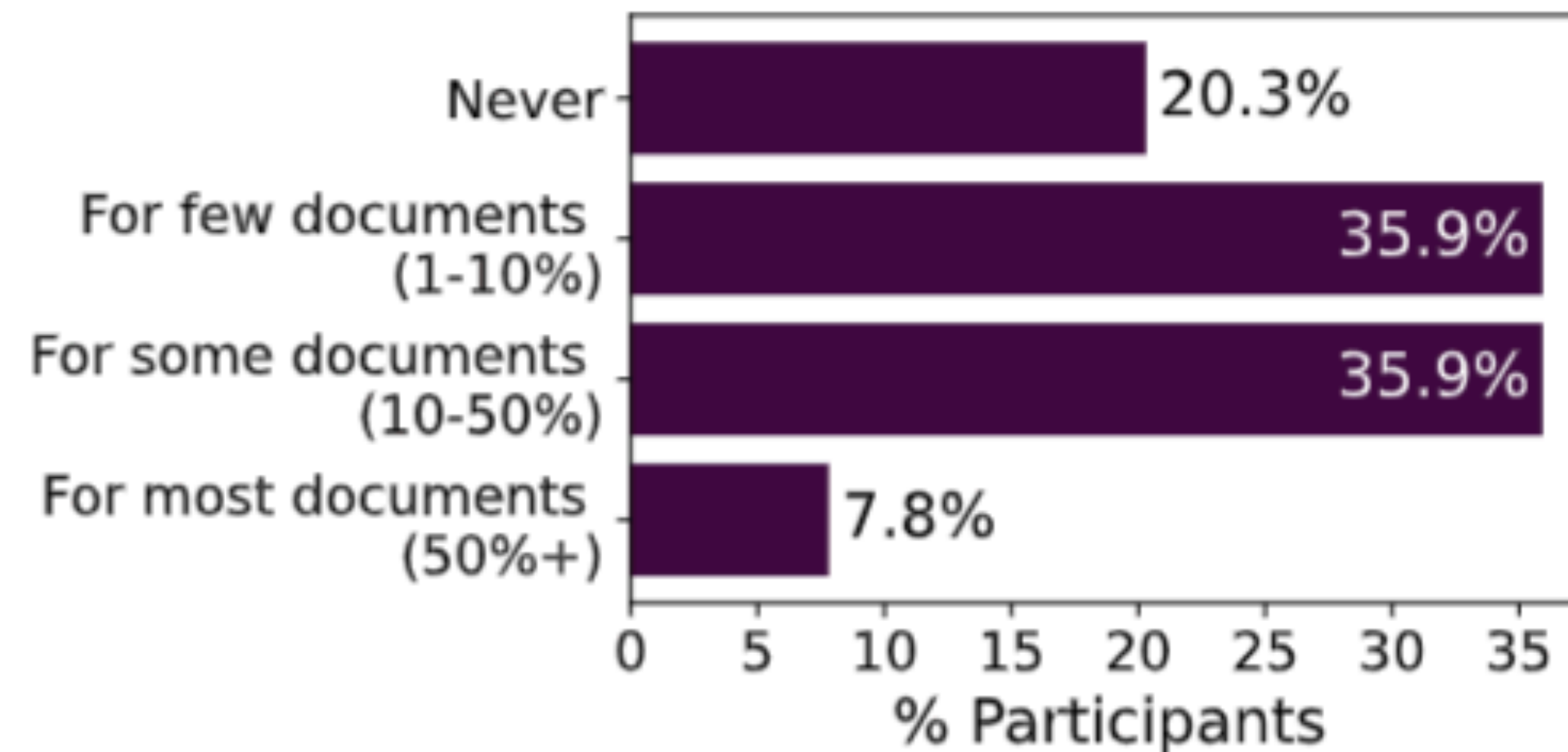


Chien-Sheng Wu



People are using chat-based LLMs for editing

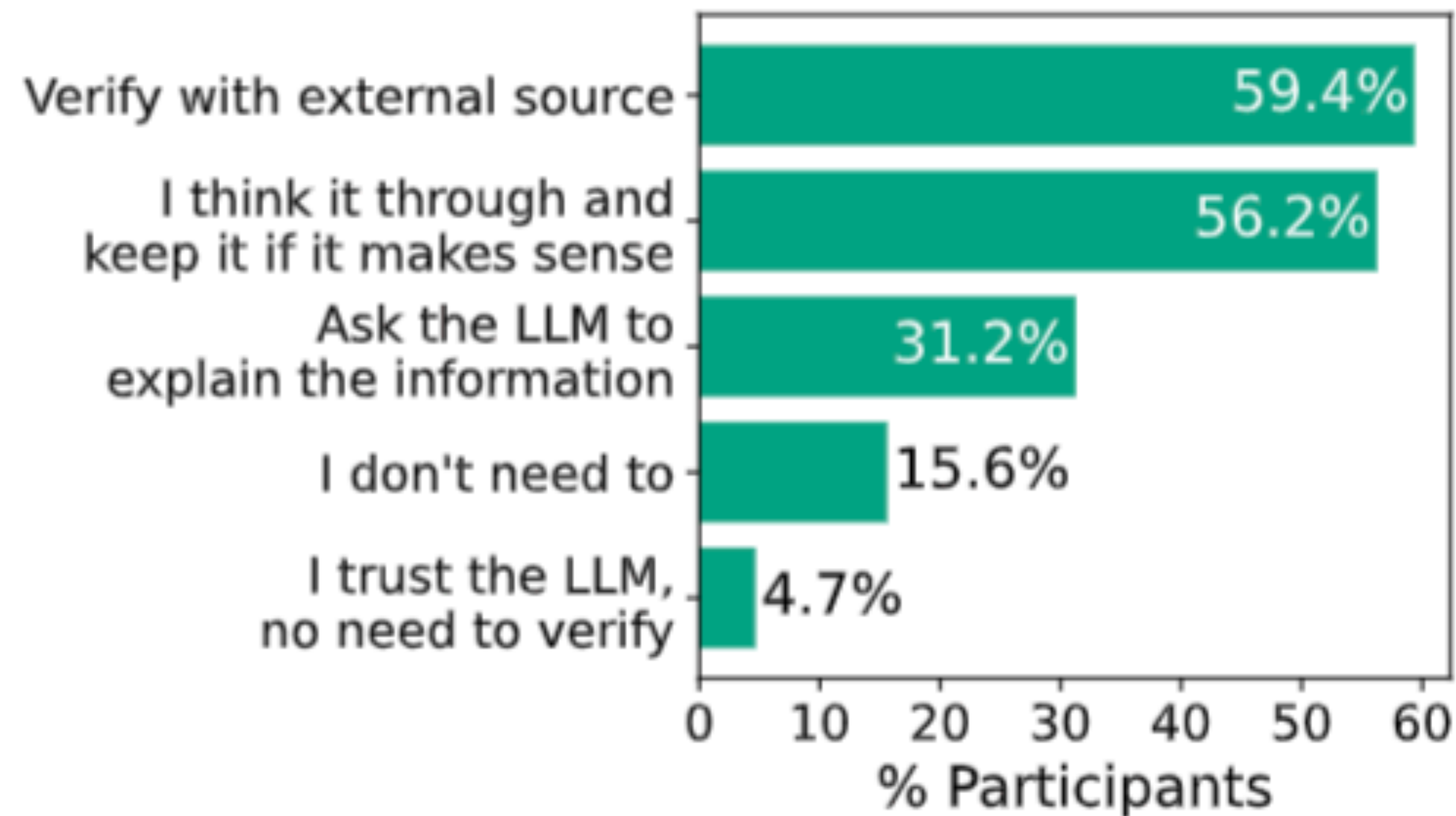
How often do you use a chat-based LLM for help in writing?



Survey of 64 people working in a large tech firm, Sept 2023

39% Technical, 29% Business, 18% sales, 14% Technical writing

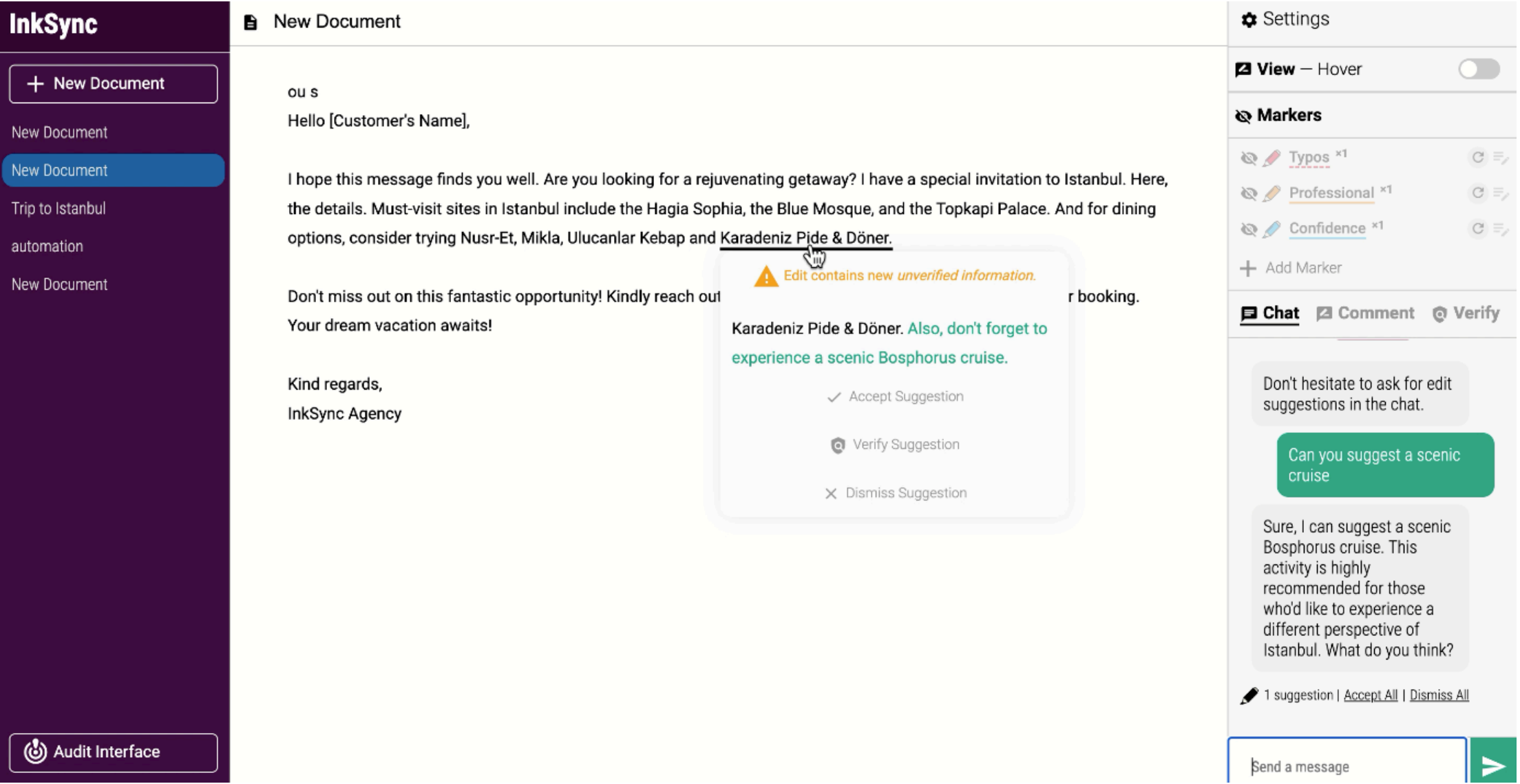
And not necessarily verifying the results...



(f) How do you verify LLM-introduced information?

Designing Conversations for Interactive Writing

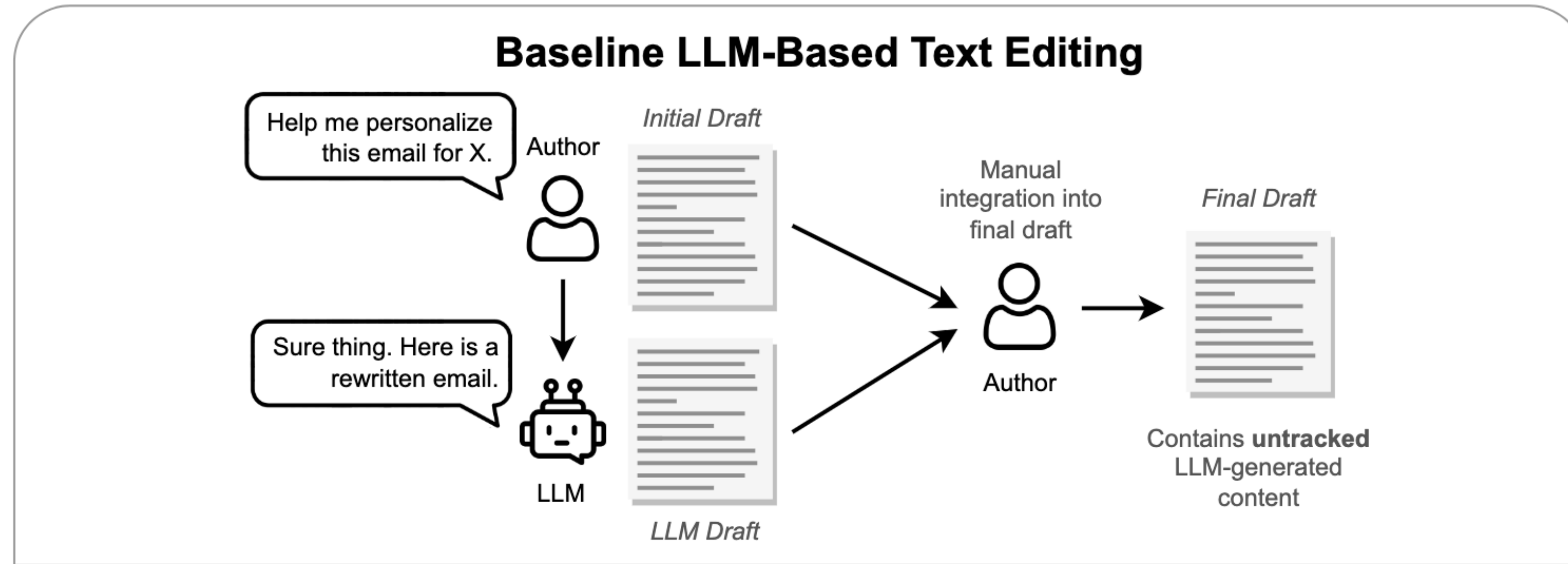
InkSync



Goals:

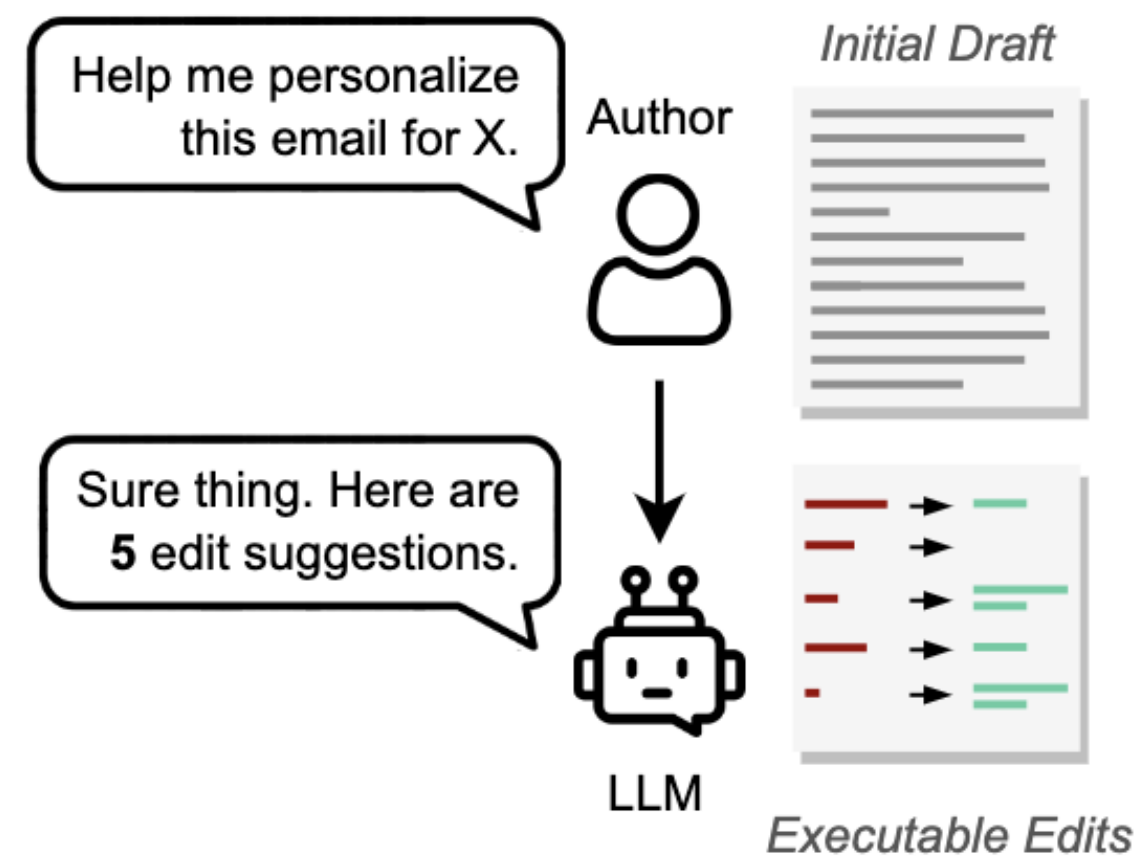
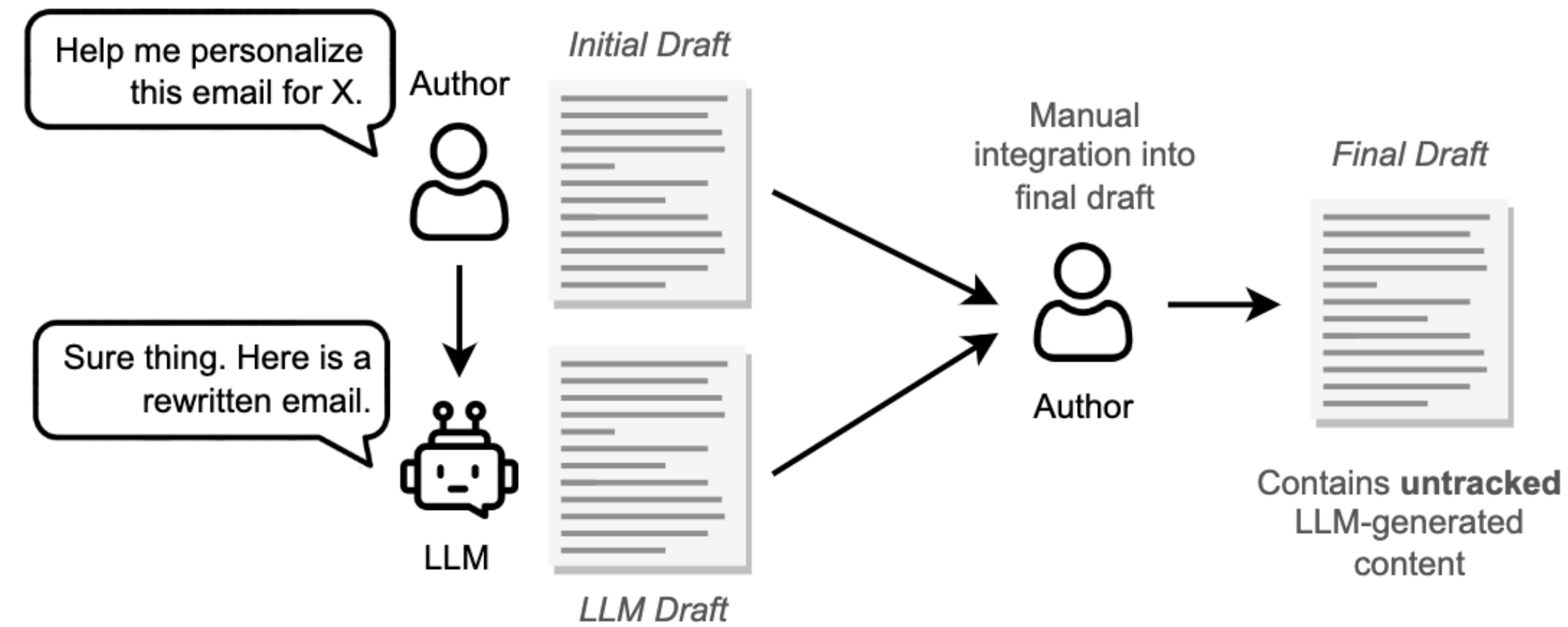
Give authors precise control over editing

Help authors / editors verify content



Standard LLM conversational interfaces lack desirable properties for text editing.

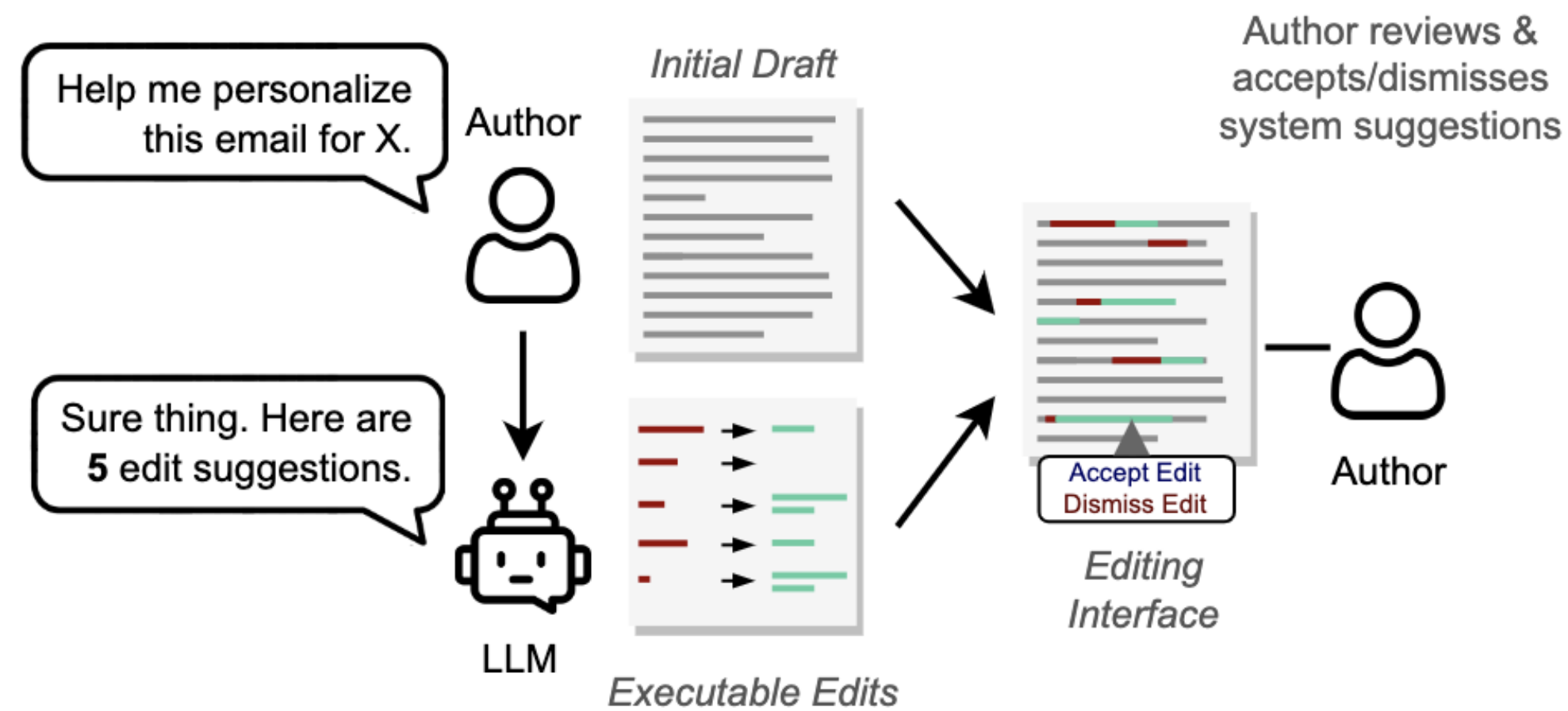
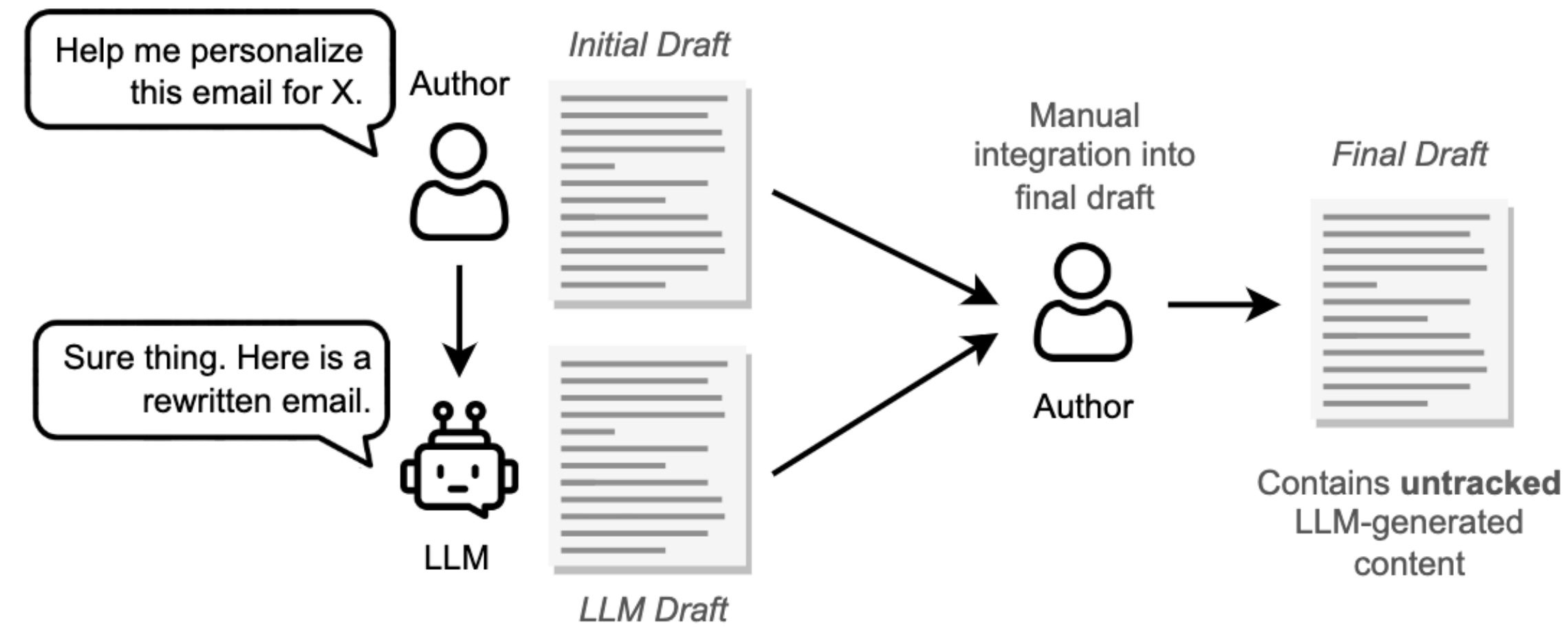
Baseline LLM-Based Text Editing



InkSync LLM-Based Text Editing

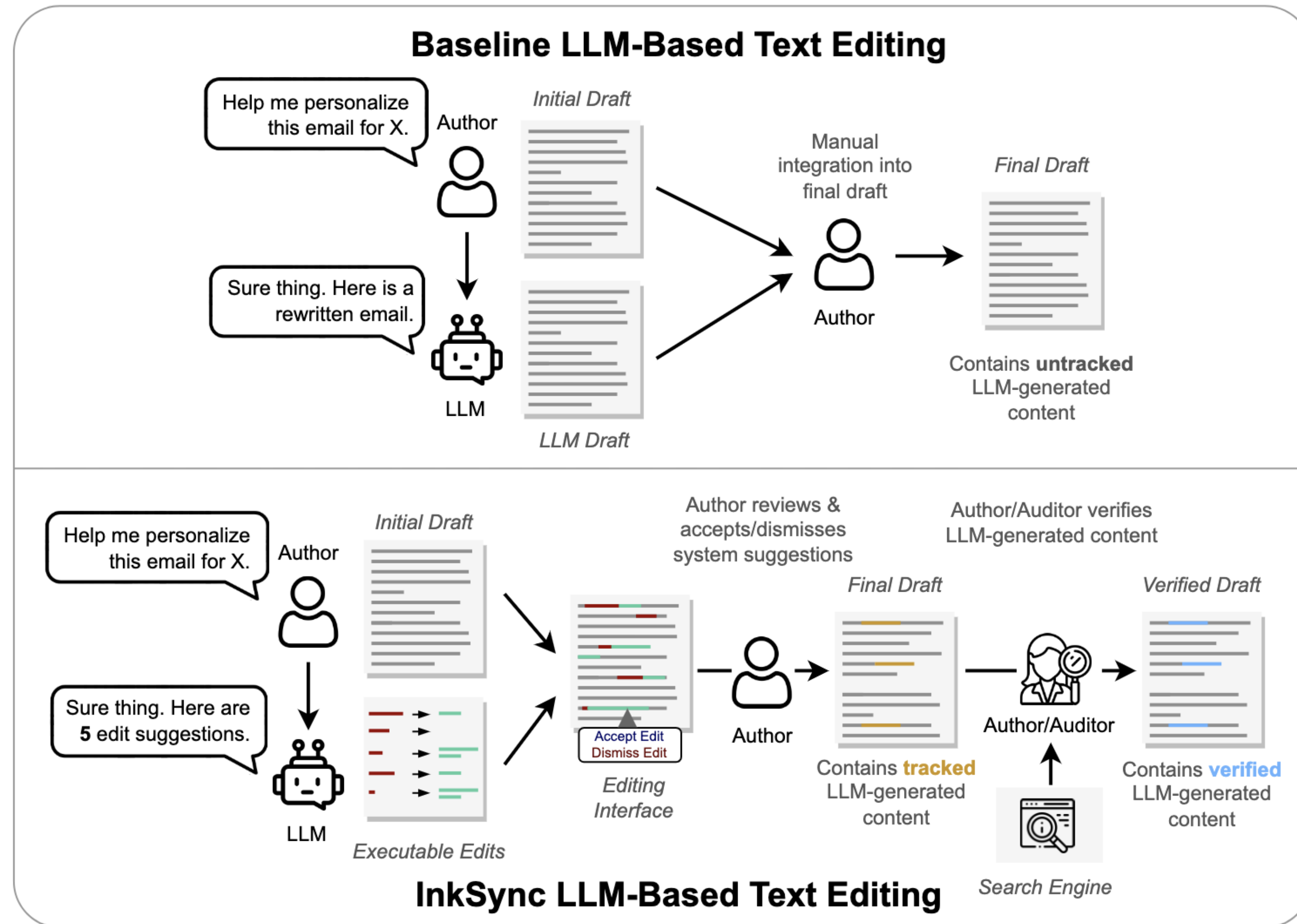
In InkSync, the LLM is instructed to generate a list of executable edits.

Baseline LLM-Based Text Editing



InkSync LLM-Based Text Editing

The author can review edit suggestions, and accept/dismiss them with one click.



A separate auditor can then verify the factual accuracy of system-generated content.

How InkSync Works:

Prompt returns executable JSON

Subject: Plan your next vacation.

Hello [Customer's Name],

I hope this message finds you well. Are you tired and want to go on a vacation? I have special invitation to Istanbul, here are the details.

Experience vibrant culture and explore the stunning landscapes that [City Destination 2] has the to offer.

Dont miss out on this fantastic opportunity! Contact us now to book your escape. Your dream vacation awaits!

Kind regards,

InkSync Agency

Don't hesitate get help in the chat.

Where is a good place to get dinner near U Chicago?

If you're looking for a dinner recommendation near U Chicago, The Promontory comes highly recommended. It offers a great variety of dishes and has a nice atmosphere. Does that help?

What kind of cuisine?

The Promontory offers a mix of cuisines, with emphasis on American comfort dishes made from locally sourced ingredients. They also have a great beverage selection. Does that help?

Send a message



How InkSync Works:

Prompt returns executable JSON

Subject: Plan your next vacation.

Hello [Customer's Name],

I hope this message finds you well. Are you planning a trip to Istanbul, here are the details.

Experience vibrant culture and explore the city's rich history.

Don't miss out on this fantastic opportunity.

Kind regards,
InkSync Agency

Executable Edit Language

```
{  
  "original_text": "Cheers,",  
  "replace_text": "Kind Regards,",  
  "component": "chat",  
  "replace_all": "0",  
  "new_info": "0"  
}
```

tion to

offer.

ion awaits!

Don't hesitate get help in the chat.

Where is a good place to get dinner near U Chicago?

If you're looking for a dinner recommendation near U Chicago, The Promontory comes highly recommended. It offers a great variety of dishes and has a nice atmosphere. Does that help?

What kind of cuisine?

The Promontory offers a mix of cuisines, with emphasis on American comfort dishes made from locally sourced ingredients. They also have a great beverage selection. Does that help?

Send a message



How InkSync Works:

UI Shows which text contains NEW INFORMATION

Subject: Plan your next vacation.

Hello [Customer's Name],

I hope this message finds you well. Are you in Istanbul, here are the details.

Experience vibrant culture and explore the city.

Dont miss out on this fantastic opportunity.

Kind regards,
InkSync Agency

historic sites like Hagia Sophia and Topkapi Palace



Edit contains new *unverified information*.

historic sites like Hagia ~~Sophia~~ Sophia,
Topkapi Palace and ~~Topkapi Palace~~, the
Istanbul Archaeology Museums,

✓ Accept Suggestion

🔍 Verify Suggestion

✗ Dismiss Suggestion

on to

fer.

on awaits!

Don't hesitate get help in the chat.

Where is a good place to get dinner near U Chicago?

If you're looking for a dinner recommendation near U Chicago, The Promontory comes highly recommended. It offers a great variety of dishes and has a nice atmosphere. Does that help?

What kind of cuisine?

The Promontory offers a mix of cuisines, with emphasis on American comfort dishes made from locally sourced ingredients. They also have a great beverage selection. Does that help?

Send a message



Editing / Verifying with InkSync Chat

InkSync

+ New Document

New Document

New Document

Trip to Istanbul

automation

New Document

 **Audit Interface**

 New Document

ou s

Hello [Customer's Name],

I hope this message finds you well. Are you looking for a rejuvenating getaway? I have a special invitation to Istanbul. Here, the details. Must-visit sites in Istanbul include the Hagia Sophia, the Blue Mosque, and the Topkapi Palace. And for dining options, consider trying Nusr-Et, Mikla, Ulucanlar Kebap and Karadeniz Pide & Döner.

Don't miss out on this fantastic opportunity! Kindly reach out to us at your earliest convenience to secure your booking.
Your dream vacation awaits!

Kind regards,
InkSync Agency

⚙ Settings

 View — Hover

Markers

  Typos x1

Professional x1

 Confidence x1

+ Add Marker

[Chat](#) [Comment](#) [Verify](#)

 Clear

Don't hesitate to ask for edit suggestions in the chat.

Send a message

How InkSync Works:

Audit Interface allows for 3rd Party Checking

New Document

— Audit Mode

Subject: Your Next Unforgettable Getaway Awaits in Malaysia!

Dear Gloria,

I trust this message finds you well. Are you looking for a vacation? Allow us to present an enticing solution – a remarkable opportunity to embark on a journey to Malaysia, where you can experience the thrill of Thai boxing, the national sport.

Envision yourself wandering through charming streets, engaging in local traditions, and uncovering hidden gems like the National Museum (Muzium Negara), along with diverse and flavorful Malaysian cuisine, particularly dishes like Nasi Goreng and Laksa that will undoubtedly evoke envy among your acquaintances back home. I encourage you to promptly reach out to us, ensuring your reservation for an extraordinary vacation that will linger in your memories indefinitely. Don't miss the chance to see the majestic Angkor Wat near Kuala Lumpur and experience the enchanting beauty of Ha Long Bay in Langkawi.

With warm regards,

InkSync Agency

Settings

Accepted Suggestions

Suggestion fbdd

uncovering hidden gems like serene museums rich with history

Suggestion ac44

uncovering hidden gems like serene museums-rich museums, along with history diverse and flavorful...

Suggestion 8ecf

Malaysian cuisine cuisine, particularly dishes like Nasi Goreng and Laksa

Suggestion ac49

serene-museums; the National Museum (Muzium Negara),

Verify

You are auditing a document that includes LLM-generated text, tagged as either:

- With unverified new information
- With no new information

Verify edits, and mark them as:

- Verified (information is correct)
- Incorrect (information is inaccurate)
- Not Sure (verification did not provide

Auditing with InkSync

InkSync

+ New Document

automation

Trip to Istanbul

New Document

Audit Interface

Trip to Istanbul

Subject: Plan your next vacation.

Hello [Customer's Name],

I hope this message finds you well. Are you looking for a rejuvenating getaway? I have a special invitation to Istanbul. Here, the details: You can explore historic sites like Hagia Sophia and Topkapi Palace, shop at the Grand Bazaar, and savor the exquisite Turkish cuisine at renowned restaurants like Nusr-Et and Çiya Sofrası.

Experience rich cultural heritage and explore the stunning landscapes such as the Bosphorus Strait and the beautiful Prince Islands that Istanbul has to offer.

Don't miss out on this fantastic opportunity! Kindly reach out to us at your earliest convenience to secure your booking. Your dream vacation awaits!

Kind regards,
InkSync Agency

Settings

View — Hover

Markers

Typos

Professional ^{x1}

Confidence

+ Add Marker

Chat

Comment

Verify

Clear

Don't hesitate to ask for edit suggestions in the chat.

elaborate on the landscapes

Sure, I can help with that. How about adding a line that mentions some stunning landscapes like the Bosphorus Strait and the Prince Islands for Istanbul? What do you think?

Send a message

STUDY 1: SIX INTERFACES COMPARED

1. All Manual

Like Google Doc without Grammarly enabled.

Baseline.

2. Markers Only

Like Google Doc with Grammarly enabled.

No chat-based component.

3. Chat Basic

Like Google Doc + a different tab with ChatGPT: a purely chat interface (no executable suggestions)

4. Chat Only

The chat can make executable suggestions anywhere into the document

5. Comments/Brainstorm

Highlight a text span, chat about how to change contents

6. InkSync Full

The full InkSync interface with markers, chat, comments & brainstorm. (2, 4, 5)

Study 1 Results

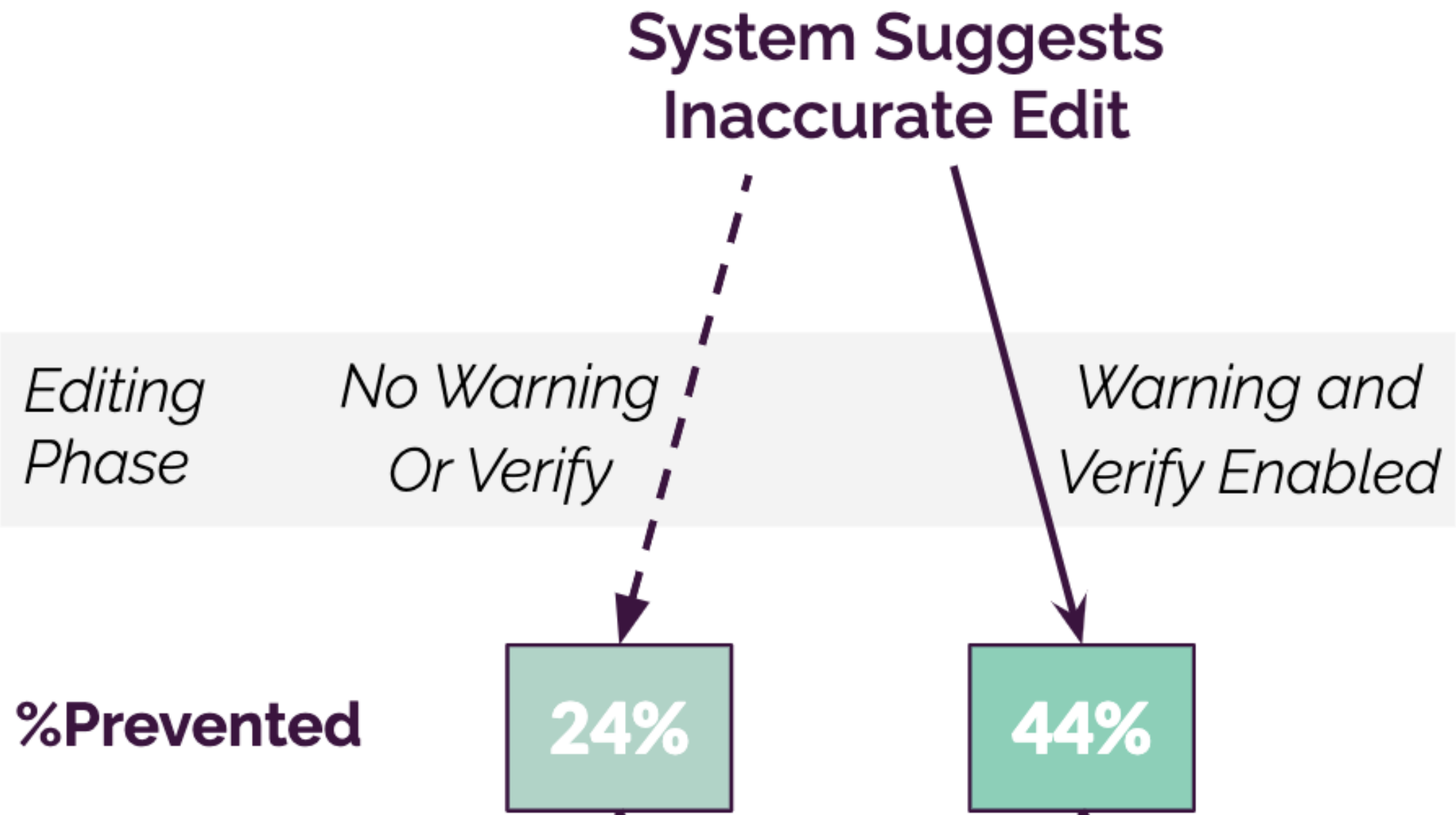
Participants:

1. Faster and more accurate with executable edits than manual
2. Preferred the structured interface
3. Felt more in control with structured chat than standard chat
4. Edits were less diverse with LLM than manual

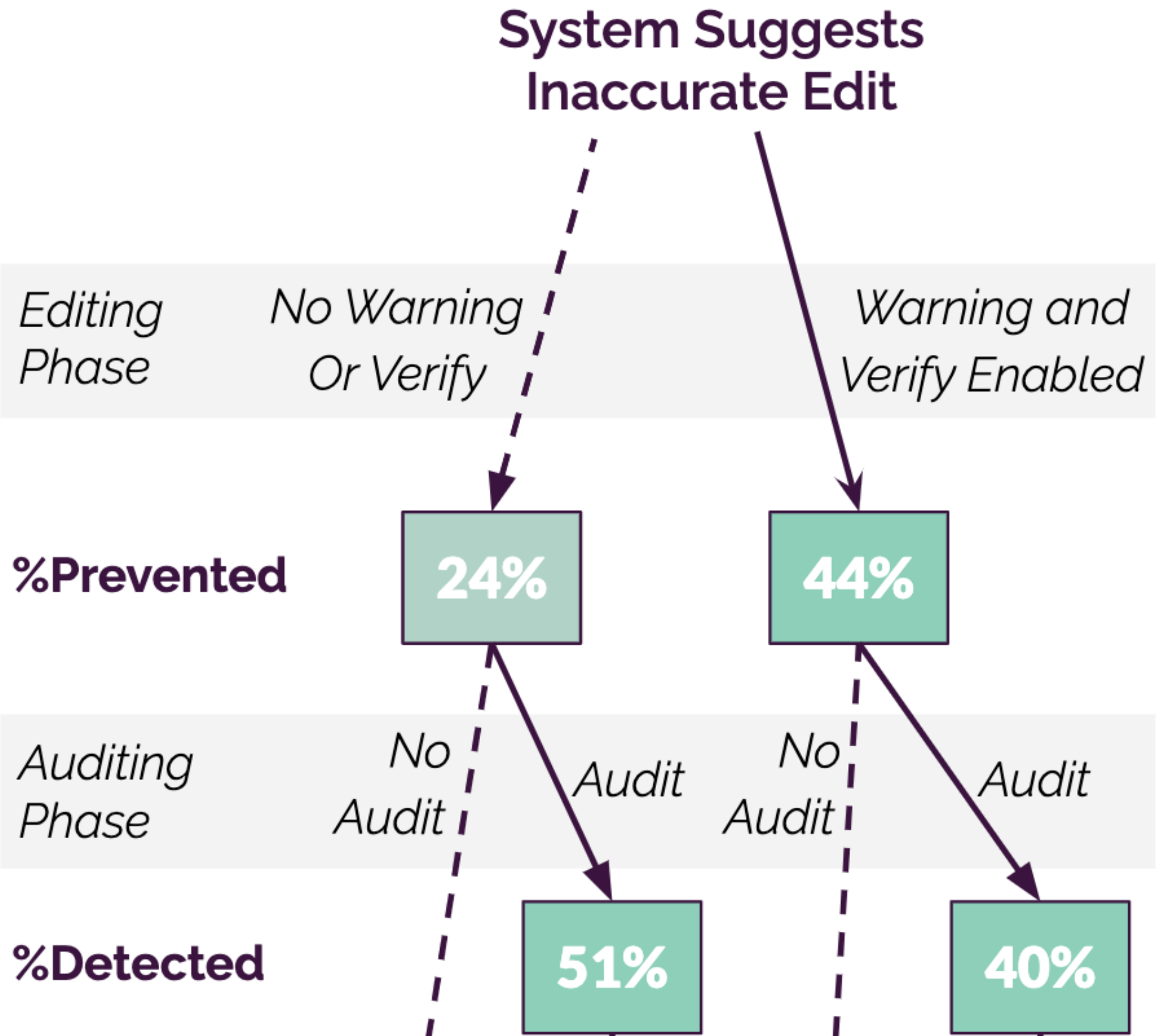
STUDY 2: ASSESS AUDIT UI

System Suggests
Inaccurate Edit

STUDY 2: ASSESS AUDIT UI



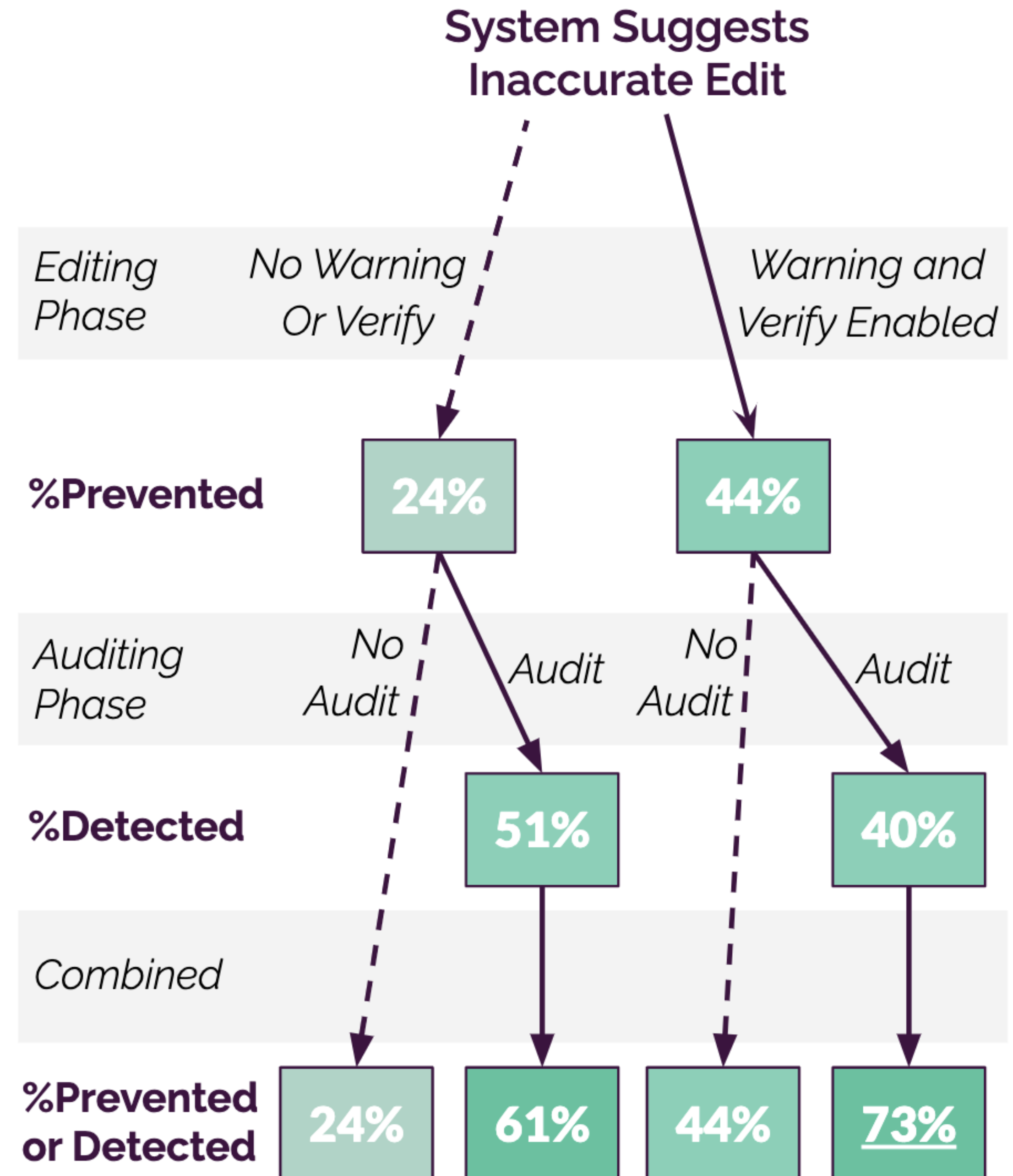
STUDY 2: ASSESS AUDIT UI



STUDY 2: ASSESS AUDIT UI

High-level findings:

1. In conditions with Warn-Verify-Audit enabled, participants prevent or detect 73% of system-generated inaccuracies.
2. That's ~3x more than in a condition without the framework.
3. Edit-time and Audit-time verification are complementary.



Summary:

Automation Bias & LLM Writing

- Automation Bias is a real concern
- Good UI can help, as shown with InkSync
 - However, have not tested this over the long run

Conclusions

Conclusions

01

Progress toward ‘Natural’ UIs is continuing steadily

02

Human-machine interaction: much remains to learn

03

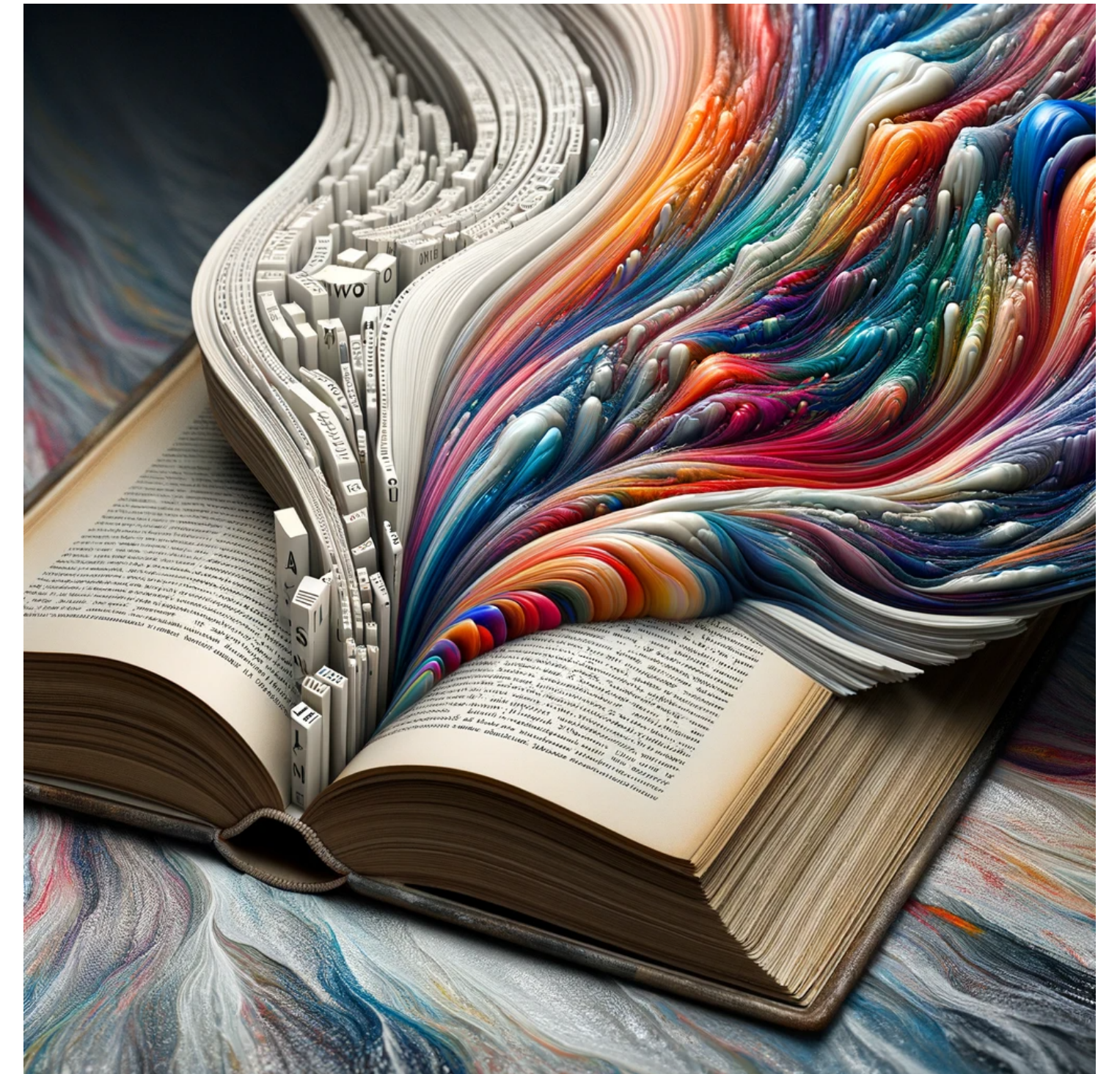
With the ubiquity of chat, automation bias is growing as a design problem in information interfaces

Conversations in LLM Apps

Thank you!

Marti Hearst
UC Berkeley

In collaboration with the Allen Institute of AI



MIDS Immersion

November 7, 2023