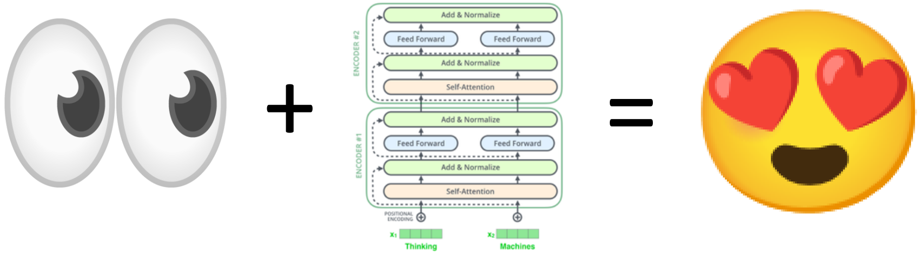
Department of Computational Linguistics

ACL 2025 Tutorial

Eye Tracking and NLP

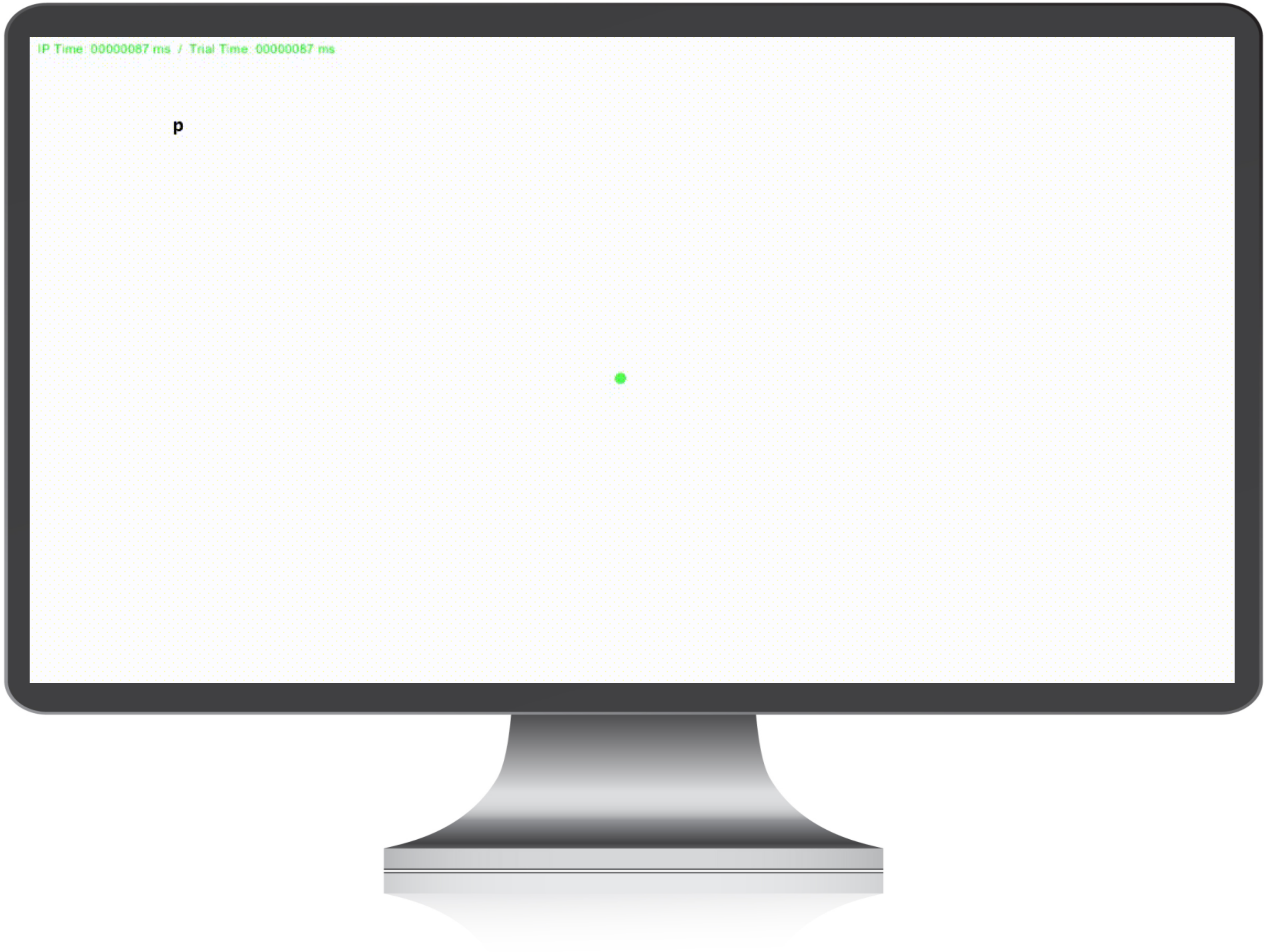


David Reich[1,2], Omer Shubi[3], Lena Jäger[1] and Yevgeni Berzak[3]

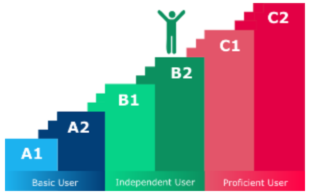
[1]University of Zurich [2]University of Potsdam [3]Technion

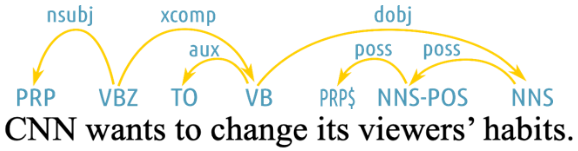
Read the Following Paragraph

Over the next 30 years, the planet's human population will increase to nine billion. Already one billion people do not get enough food. The increase will mean more pressure on agricultural land, water, forests, fisheries and biodiversity resources, as well as nutrients and energy supplies. There is also the issue of methane excreted by cows. The livestock farming contribution, in terms of greenhouse gas emissions, is enormous – 35% of the planet's methane, 65% of its nitrous oxide and 9% of the carbon dioxide.

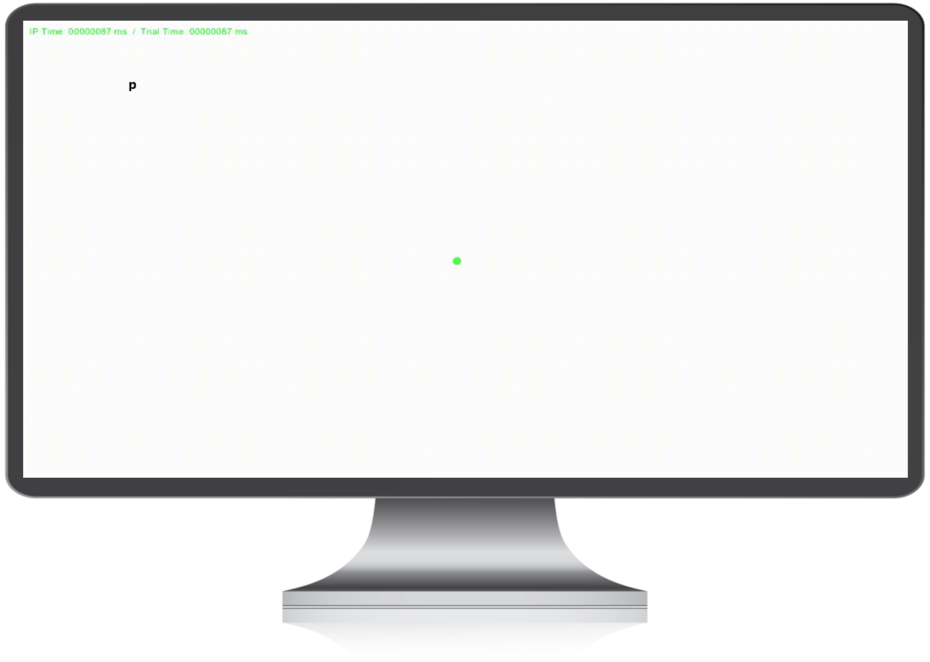

IP Time: 00000087 ms / Trial Time: 00000087 ms
p

L1

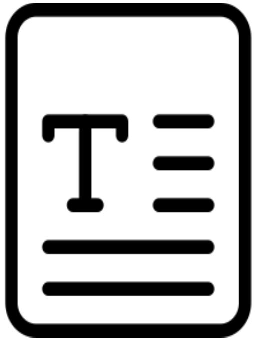

A1
A2
B1
B2
C1
C2
Basic User
Independent User
Proficient User


nsubj
xcomp
aux
dobj
poss
poss
PRP VBZ TO VB PRP$ NNS-POS NNS
CNN wants to change its viewers' habits.

Eye Movements in Reading: Rich Information



Comprehension
Relevance
Goals
Prior exposure
…



Linguistic knowledge
Reading skill
Cognitive state
…

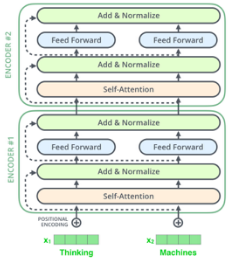
Ling. characteristics
Information structure
Difficulty level
…

This Tutorial

Reading Research

Science

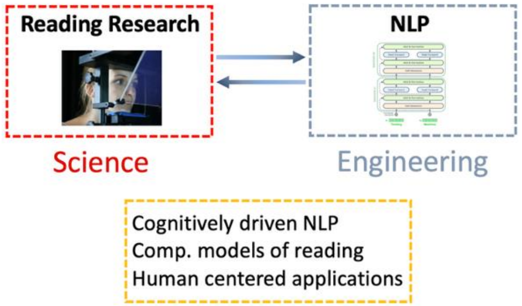
NLP



Engineering

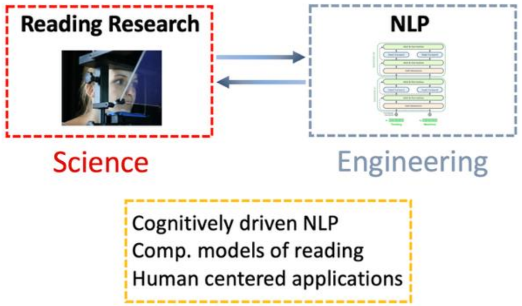
Cognitively driven NLP
Comp. models of reading
Human centered applications

Huge untapped potential for NLP research



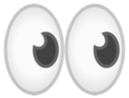
- A unique multimodal modeling challenge
- Expanding the role of NLP in cognitive modeling and science
- Opportunities for innovative high-impact applications
  - Education
  - Language learning and assessment
  - Content personalization
  - Content accessibility
  - …

The domain is ripe enough

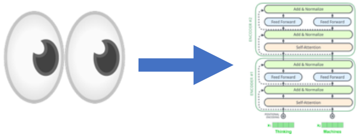


- Builds on a long tradition in the psychology of reading
- Amount and diversity of eye tracking data has reached a critical mass
- Recent NLP and ML modelling approaches demonstrate feasibility

Tutorial Outline

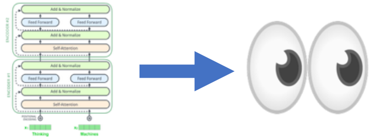


1. Introduction to eye tracking

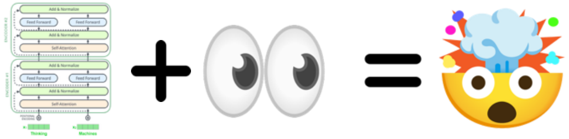


2. Uses of eye tracking in NLP + QA

30 minutes break



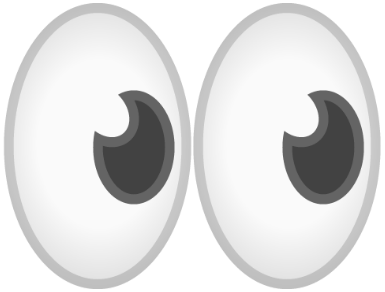
3. NLP for eye movement and cognitive modeling



4. New human centered applications

5. Outlook and future directions + QA

Introduction to Eye Movements in Reading and Eye Tracking



Introduction to Eye Movements in Reading and Eye Tracking

How do People Read?

Data Representation

Reading Measures

Alternatives

Eye Tracking

Datasets

Introduction to Eye Movements in Reading and Eye Tracking

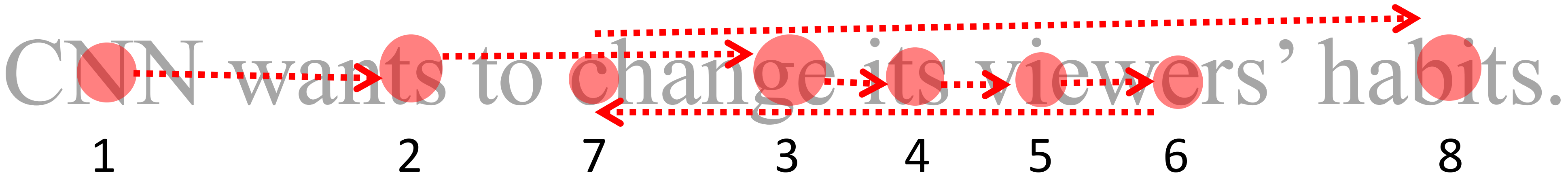
- **How do People Read?**
- Data Representation
- Reading Measures
- Alternatives
- Eye Tracking
- Datasets

Read the Following Paragraph

Over the next 30 years, the planet's human population will increase to nine billion. Already one billion people do not get enough food. The increase will mean more pressure on agricultural land, water, forests, fisheries and biodiversity resources, as well as nutrients and energy supplies. There is also the issue of methane excreted by cows. The livestock farming contribution, in terms of greenhouse gas emissions, is enormous – 35% of the planet's methane, 65% of its nitrous oxide and 9% of the carbon dioxide.

How do people read?

How do people read?



How do people read?

Fixations

How do people read?

Saccades

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

How do people read?

CNN wants to change its viewers' habits.

225ms

How do people read?

CNN wants to change its viewers' habits.

225ms

How do people read?

CNN wants to change its viewers' habits.

225ms 30ms

How do people read?

CNN wants to change its viewers' habits.

What do you see during a **fixation**?

How do people read?

CNN wants to change its viewers' habits.

Perceptual Span

What do you see during a **fixation**?

How do people read?

CNN wants to change its viewers' habits.

What do you see during a **saccade**?

How do people read?

What do you see during a **saccade**?

Nothing!

How do people read?

CNN wants to change its viewers' habits.

Forward Saccade

How do people read?

CNN wants to change its viewers' habits.

Forward Saccade

How do people read?

CNN wants to change its viewers' habits.

Forward
Saccade

How do people read?

CNN wants to change its viewers' habits.

Backward
Saccade
(Regression)

How do people read?

How do people read?

CNN wants to change its viewers' habits.

1 2 7 3 4 5 6 8

How do people read?

CNN wants to change its viewers’ habits.

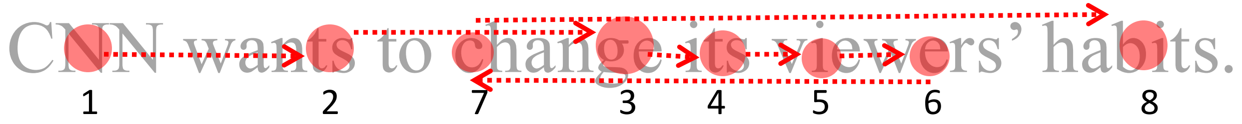
1 2 7 3 4 5 6 8

How do people read?

CNN wants to change its viewers' habits.

1 2 7 3 4 5 6 8

How do people read?



Eye Mind Assumption: "… *there is no appreciable lag between what is fixated and what is processed*." Just & Carpenter, 1980

Tight correspondence between eye movements and linguistic processing

- Eye movements capture online processing difficulty:
  e.g. longer, less frequent and less predictable words
  → longer fixation times, less skipping

Introduction to Eye Movements in Reading and Eye Tracking

How do People Read?

Data Representation

Reading Measures

Alternatives

Eye Tracking

Datasets

Reading Measures

?

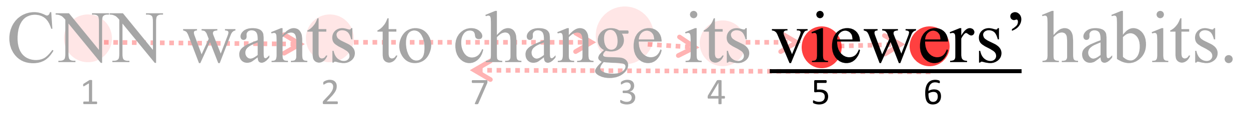
CNN wants to change its viewers' habits.

- Skips (also skip rate / fixation probability)
- First fixation duration
- Gaze duration
- Regression rate
- Go-past duration
- Total fixation duration

Reading Measures

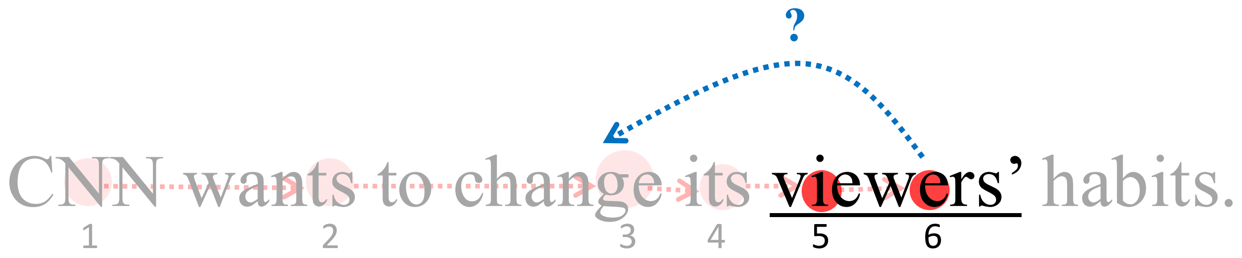
- Skips (also skip rate / fixation probability)
- First fixation duration
- Gaze duration
- Regression rate
- Go-past duration
- Total fixation duration

Reading Measures



- Skips (also skip rate / fixation probability)
- First fixation duration
- **Gaze duration**
- Regression rate
- Go-past duration
- Total fixation duration

Reading Measures



- Skips (also skip rate / fixation probability)
- First fixation duration
- Gaze duration
- Regression rate
- Go-past duration
- Total fixation duration

Reading Measures

- Skips (also skip rate / fixation probability)
- First fixation duration
- Gaze duration
- Regression rate
- Go-past duration
- Total fixation duration

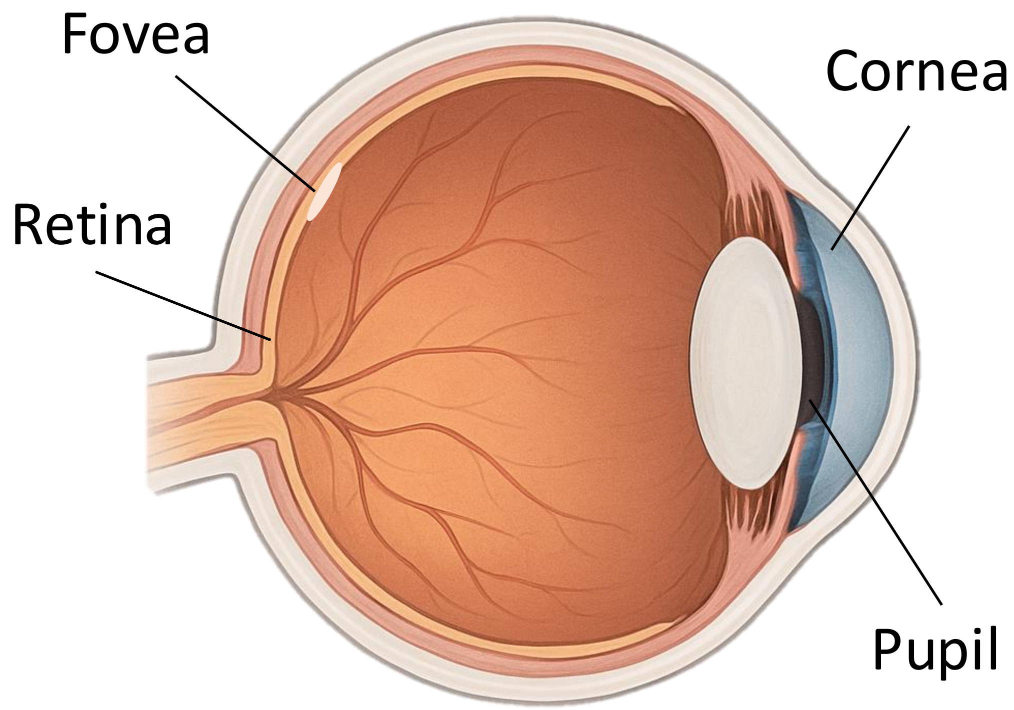
Reading Measures

- Skips (also skip rate / fixation probability)
- First fixation duration
- Gaze duration
- Regression rate
- Go-past duration
- Total fixation duration

Introduction to Eye Movements in Reading and Eye Tracking

- How do People Read?
- Data Representation
- Reading Measures
- Alternatives
- **Eye Tracking**
- Datasets

Eye Physiology



Photoreceptor Cells on the Retina

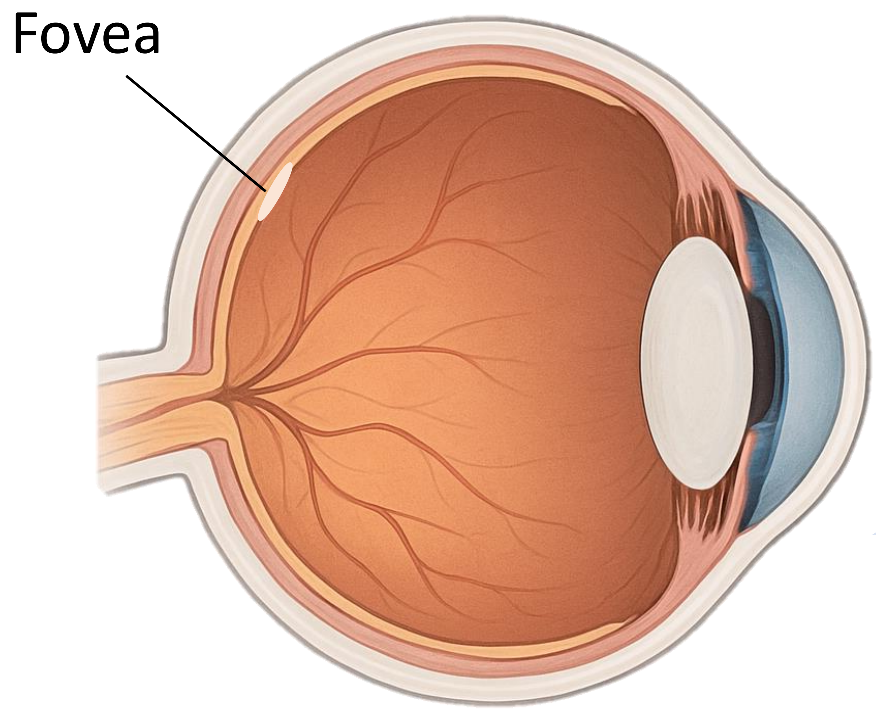
Cones

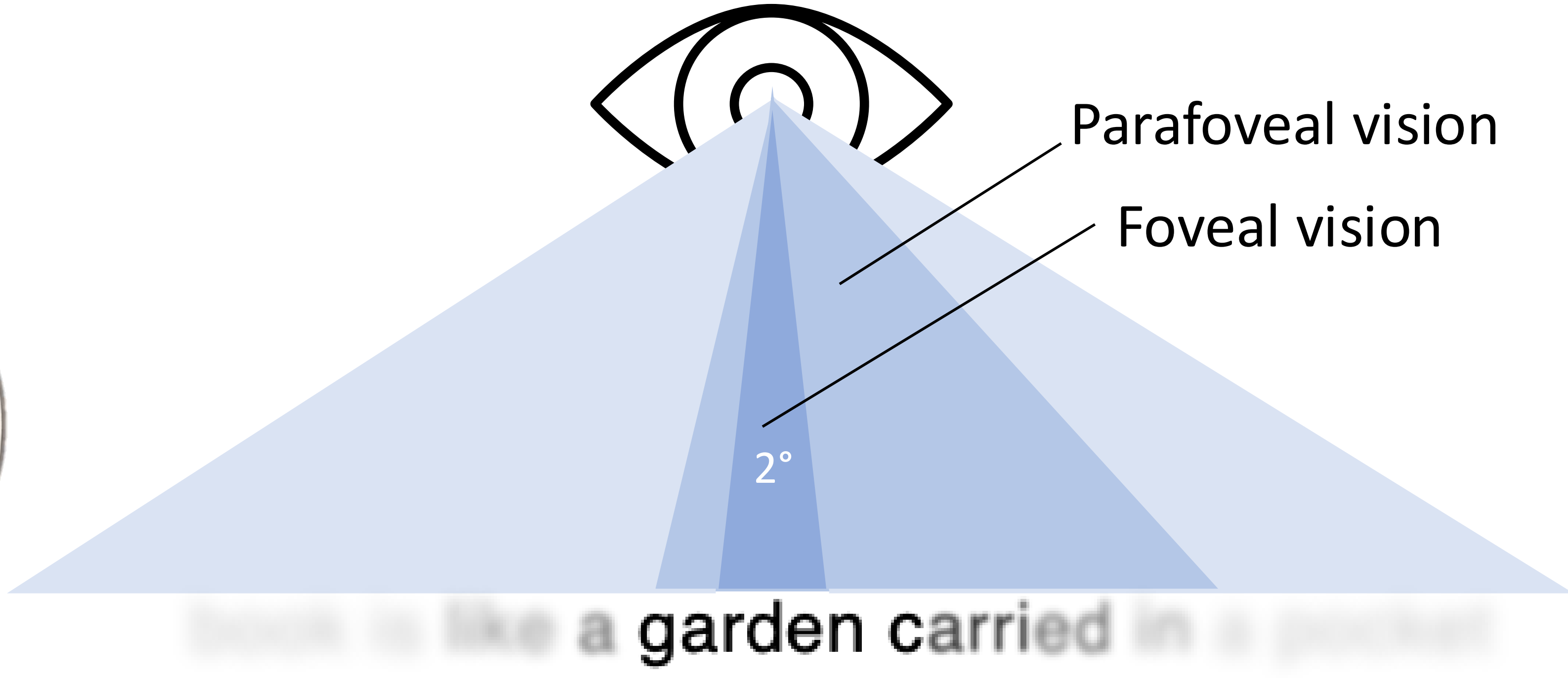
- Sensitive to “visual detail” (spatial frequency and color)
- High density in the fovea
- Low density in the peripheria

Rods

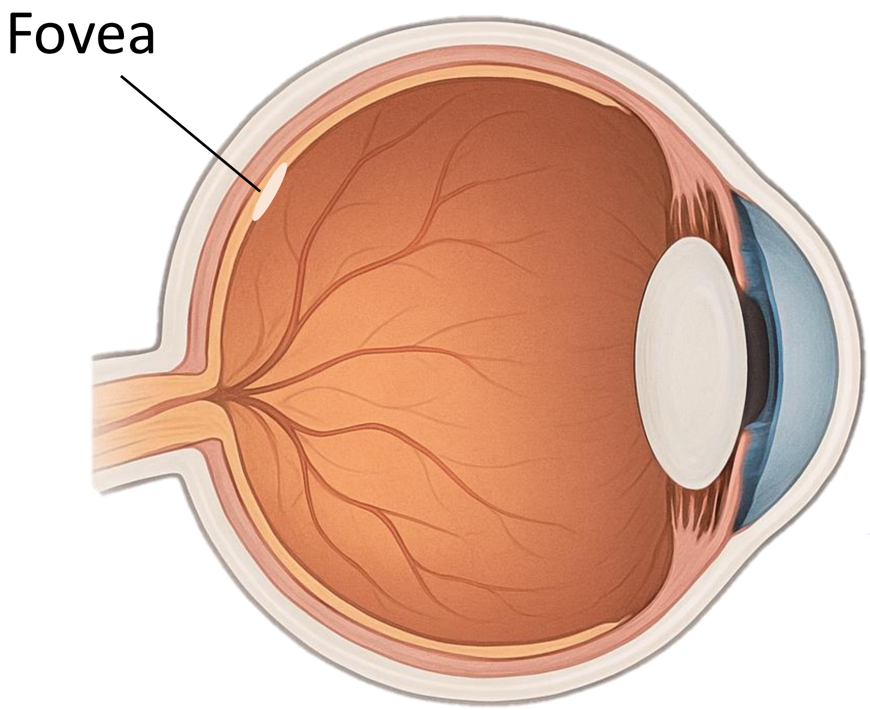
- Sensitive to light
- Low density in the fovea
- High density in the peripheria

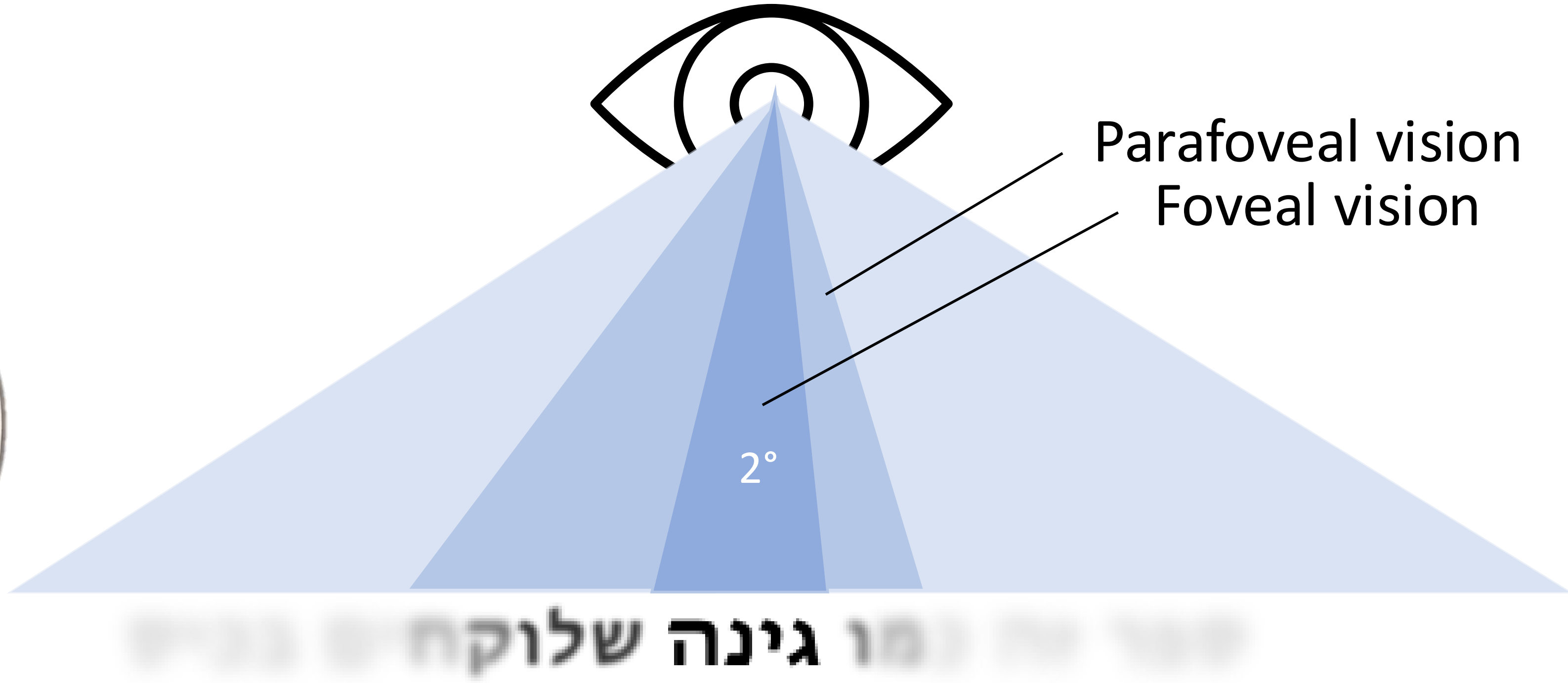
The Perceptual Span in Reading



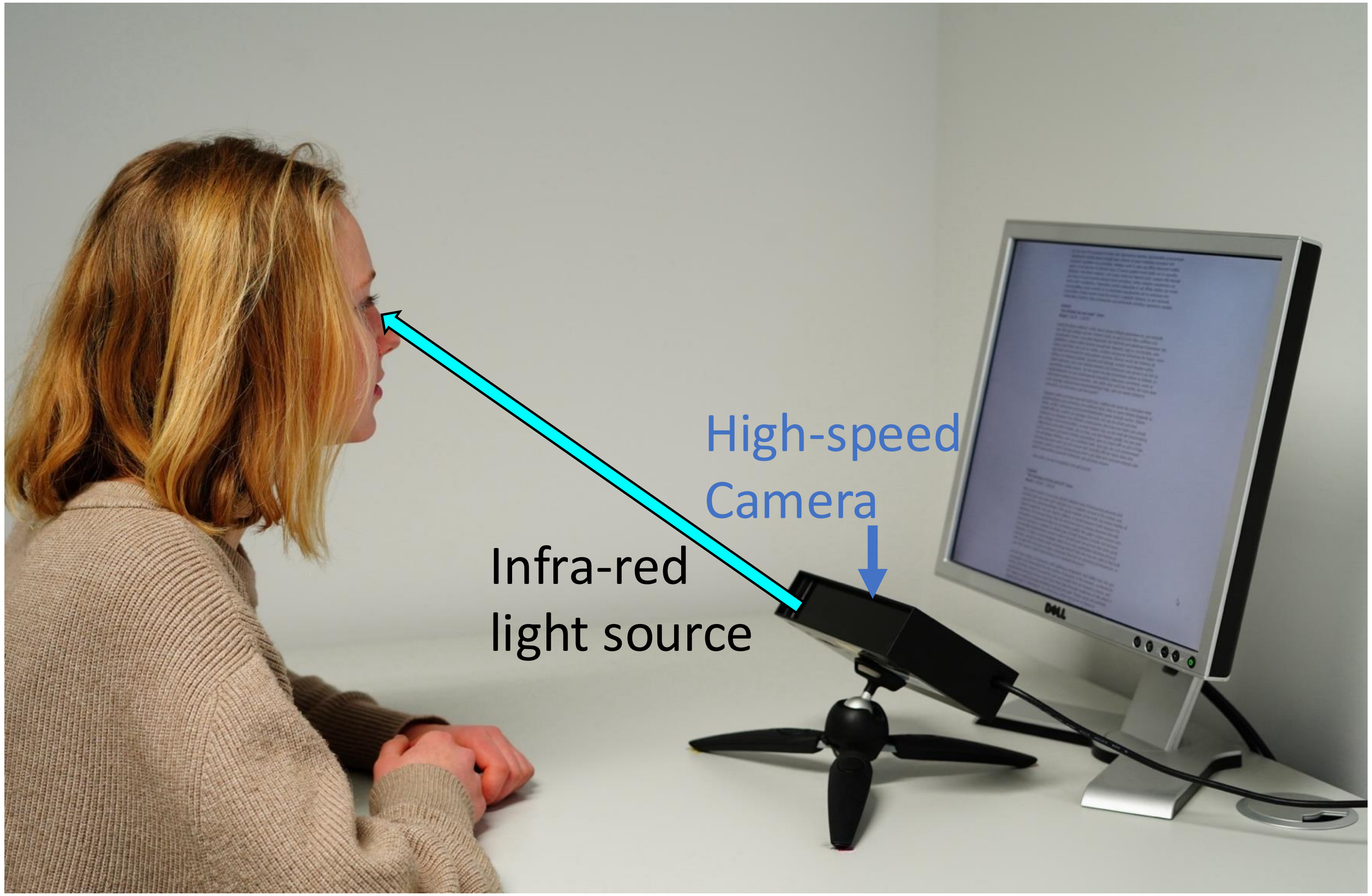


The Perceptual Span in Reading

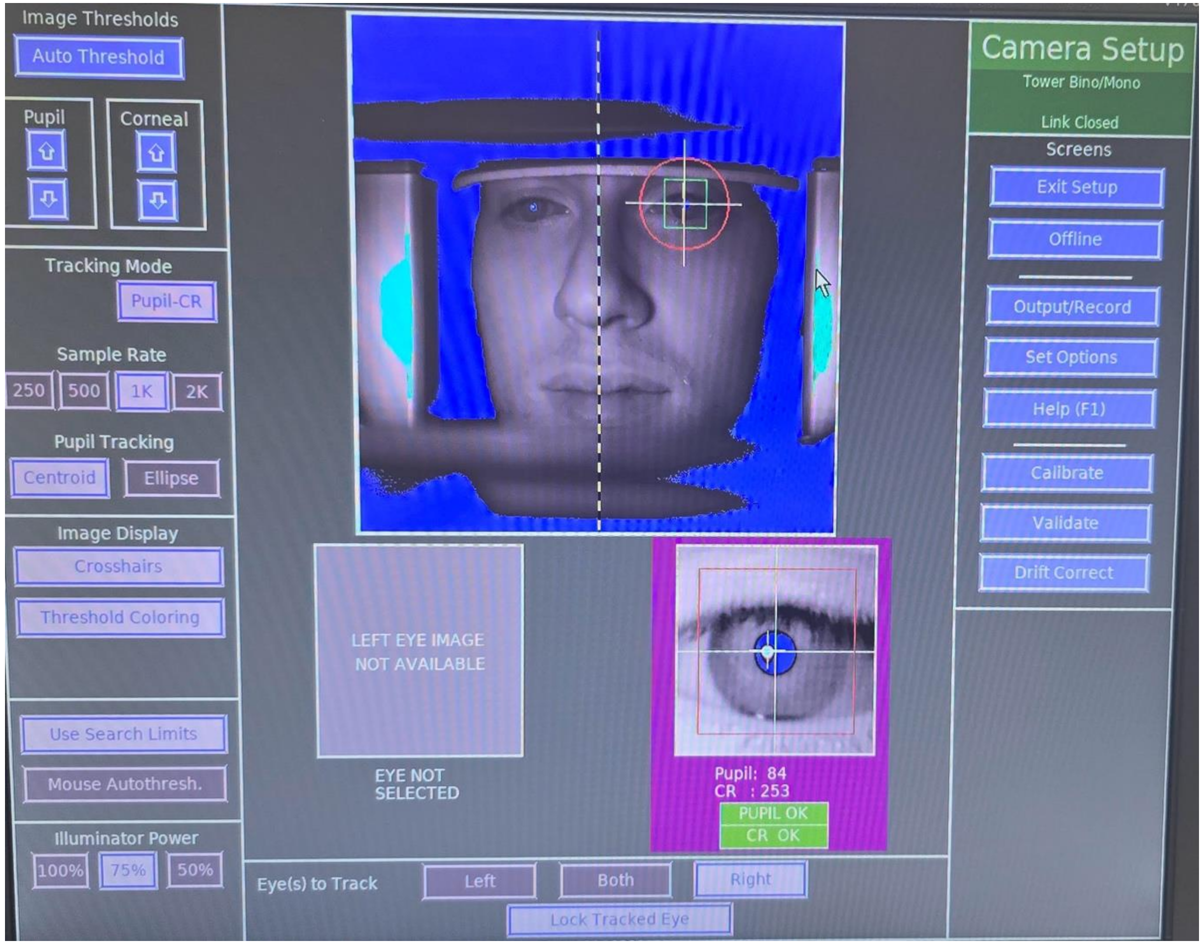




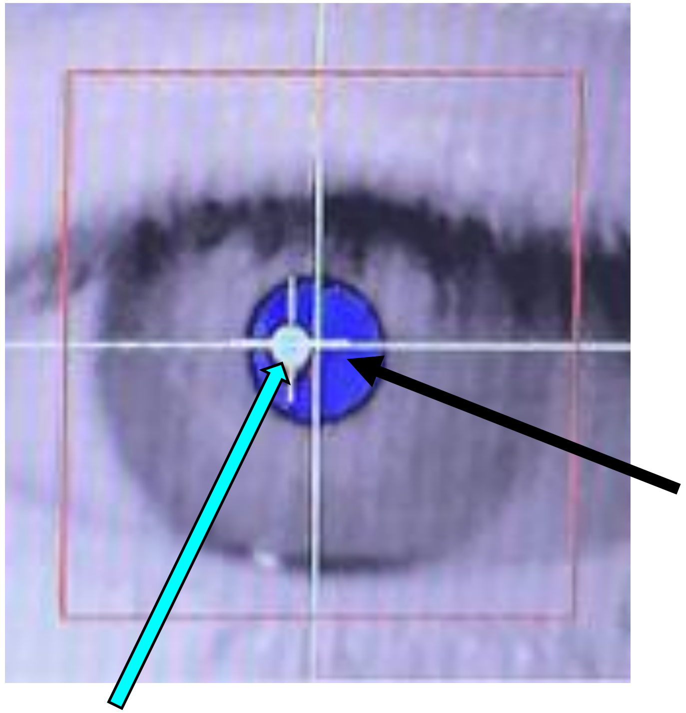
Video-Based Eye Tracking



Video-Based Eye Tracking



Two targets



Pupil

Corneal Reflection
of the infra-red light

Video-Based Eye Tracking

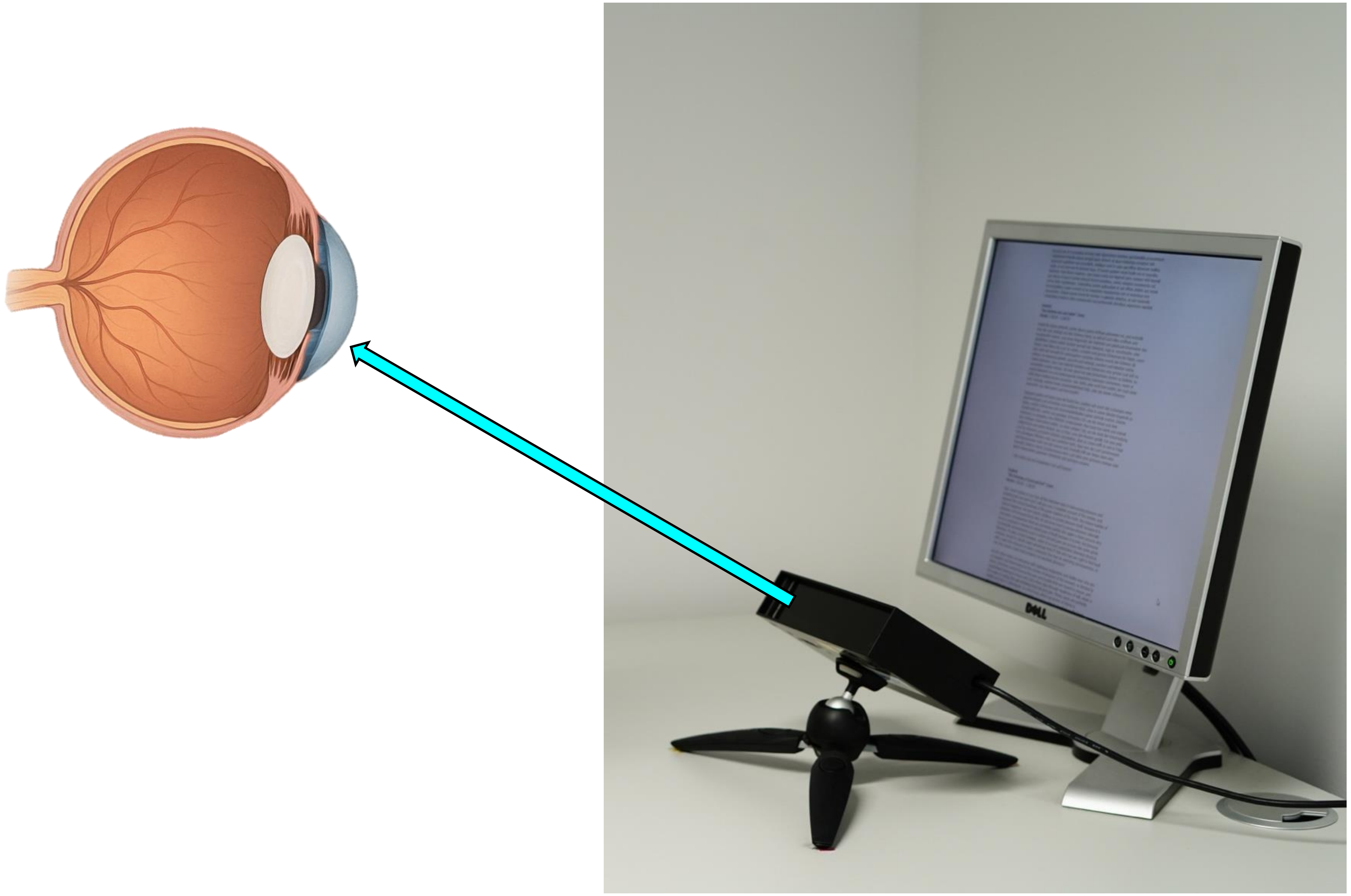
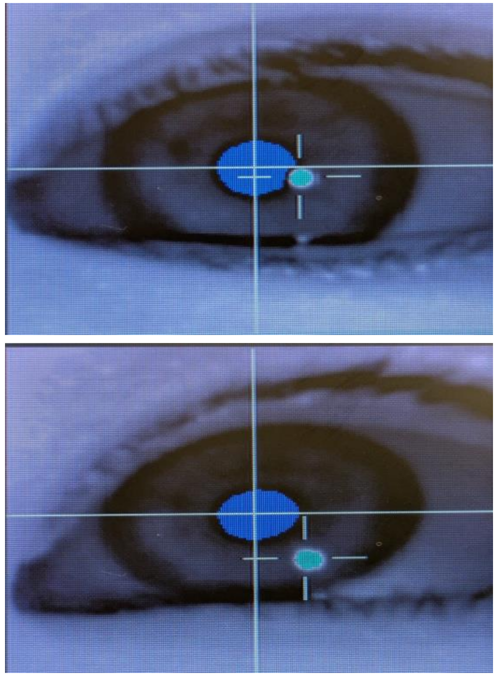


Image detection of the
- pupil
- corneal reflection (CR) of the IR light



Video-Based Eye Tracking

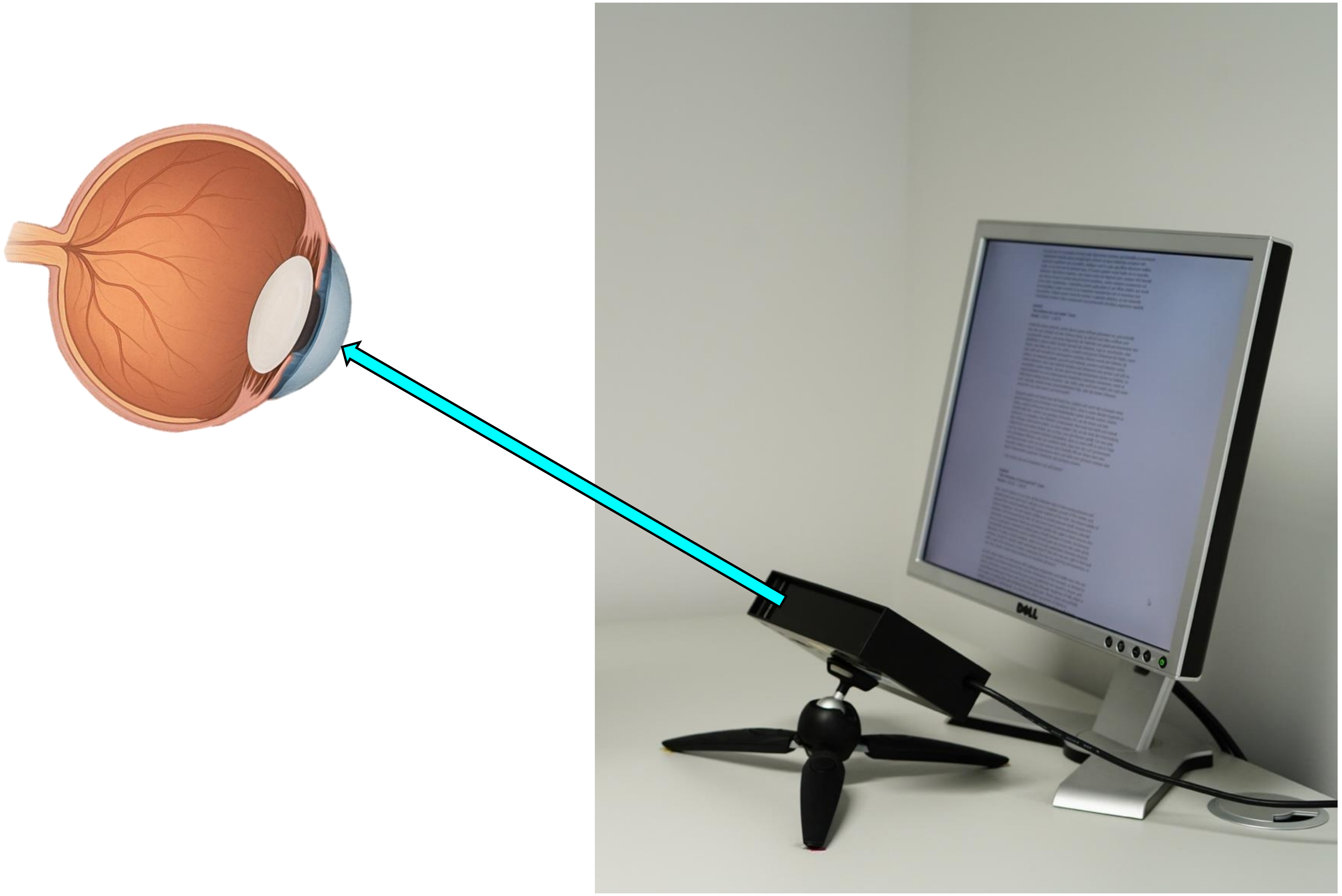
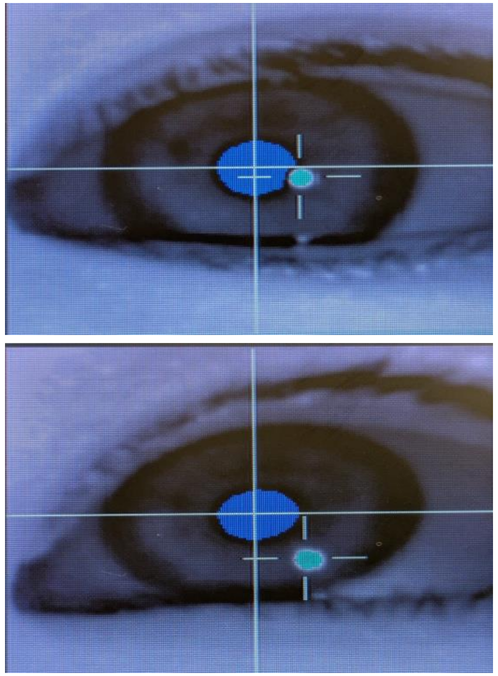
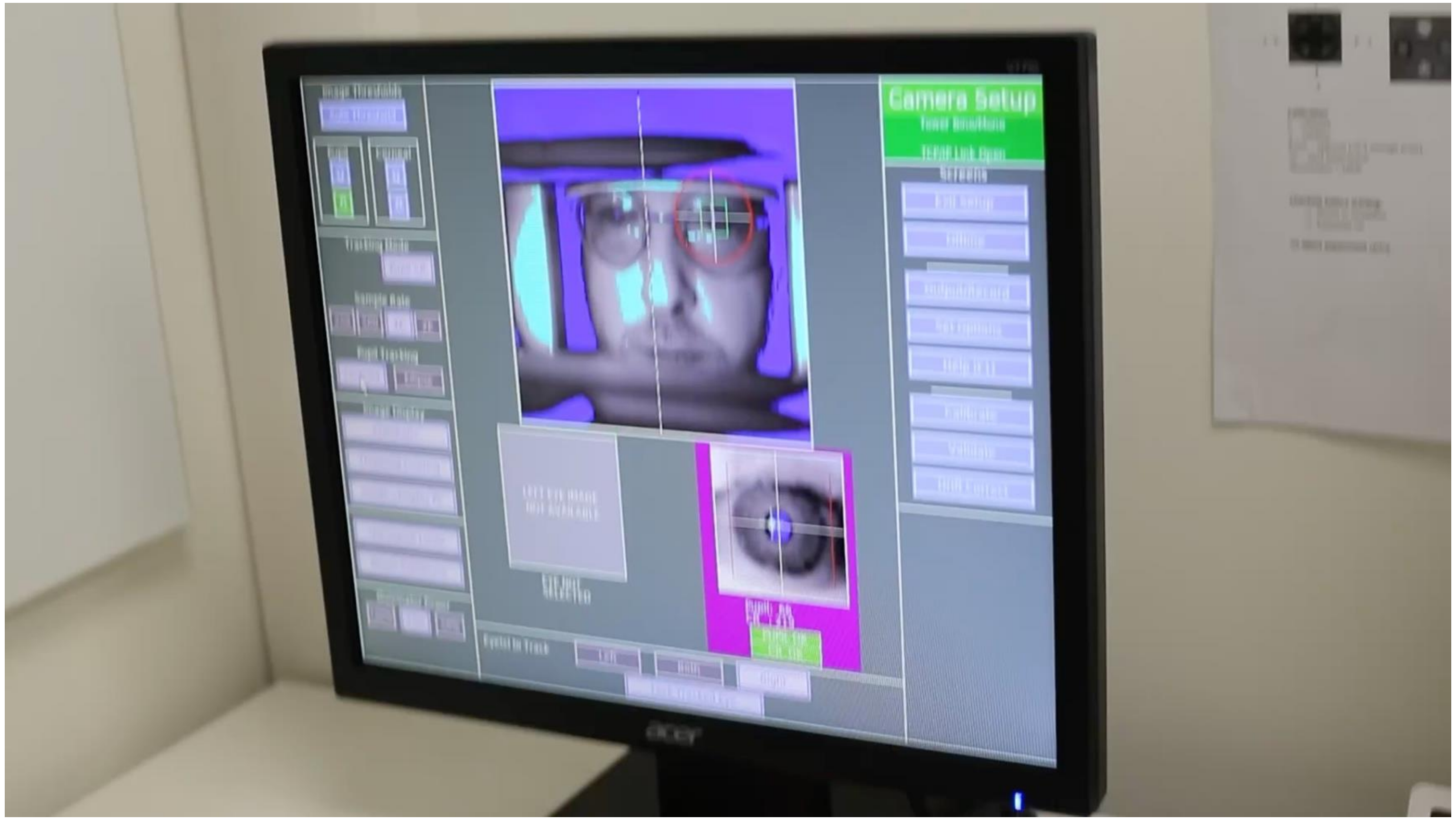


Image detection of the
- pupil
- corneal reflection (CR) of the IR light



Video-Based Eye Tracking

Calibration



Very high calibration quality needed for reading research

[Calibration Video]

Video-Based Eye Tracking Devices

Wearable

Portable

Stationary

Video-Based Eye Tracking Devices

For eye tracking-while-reading data sets we typically want

- character-level spatial resolution
  - very high accuracy (calibration quality) needed
  - head-stabilization (chin-rest) recommended
  - stationary or portable devices typically achieve better calibration than wearables
- precise fixation onset/offset times
  - sampling frequency of at least 200 Hz needed

Data Collection Considerations

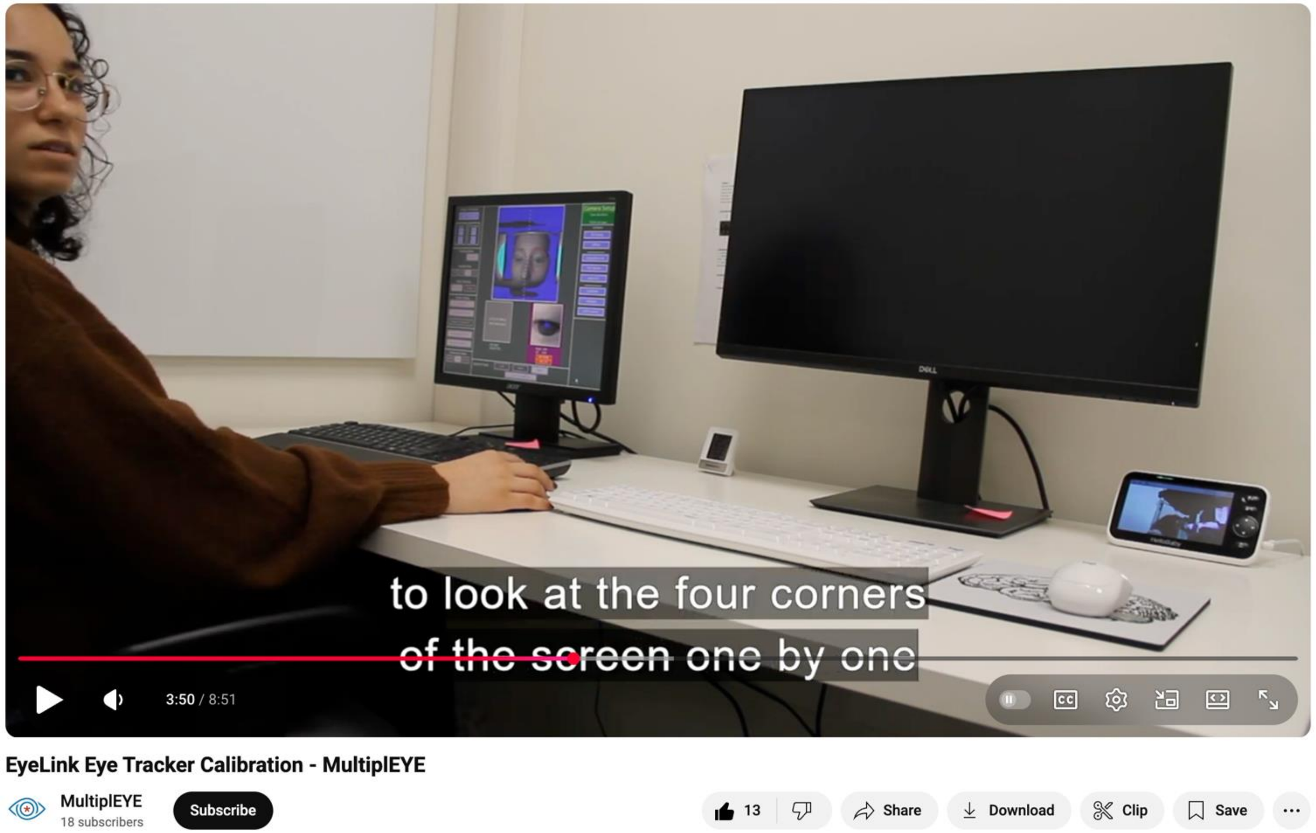
Many additional things to take care of

- Experimental design, counter balancing and randomization
- Attention checks / reading comprehension questions
- Monitoring drift and recalibration during the experiment
- Blocking accidental clickthrough’s
- Text presentation:
  - Font (often monospace) and font size
  - Line spacing
- IRB (ethics approval, data protection etc.)
- …

Data Collection Considerations

Data collection video tutorials (for EyeLink Eye Trackers)

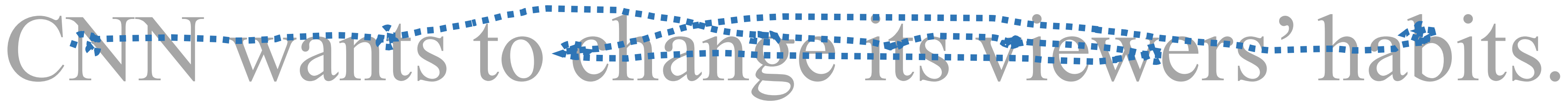
- Dominant eye test
- Calibration



Introduction to Eye Movements in Reading and Eye Tracking

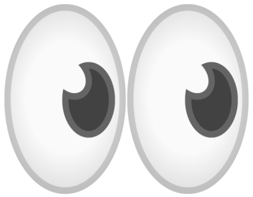
- How do People Read?
- Data Representation
- Reading Measures
- Alternatives
- Eye Tracking
- Datasets

Eye Tracking: Recorded data



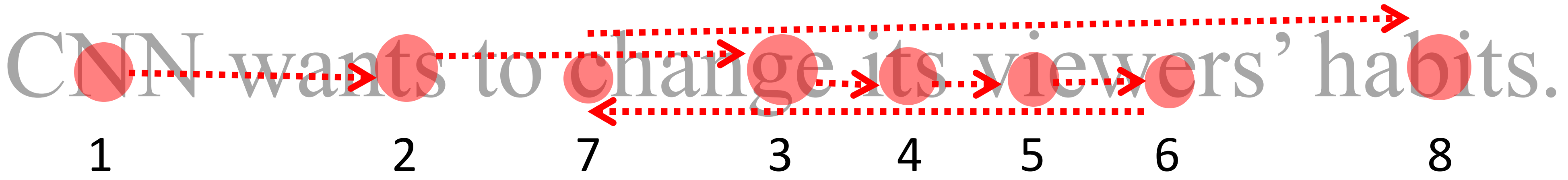
Raw Data

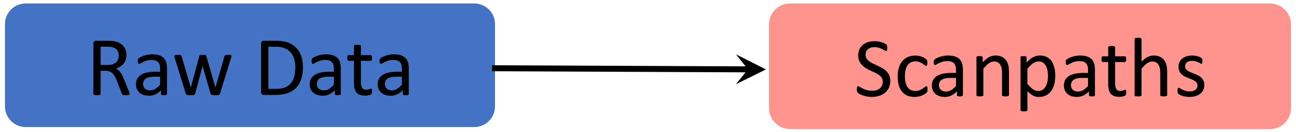
- Binocular or monocular



- Screen coordinates or visual angle

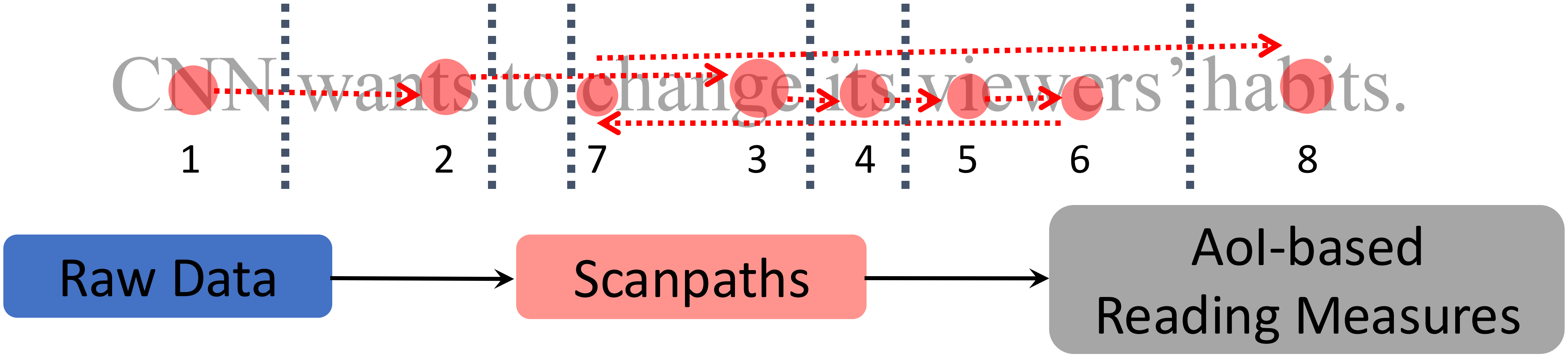
Eye Tracking: Preprocessing





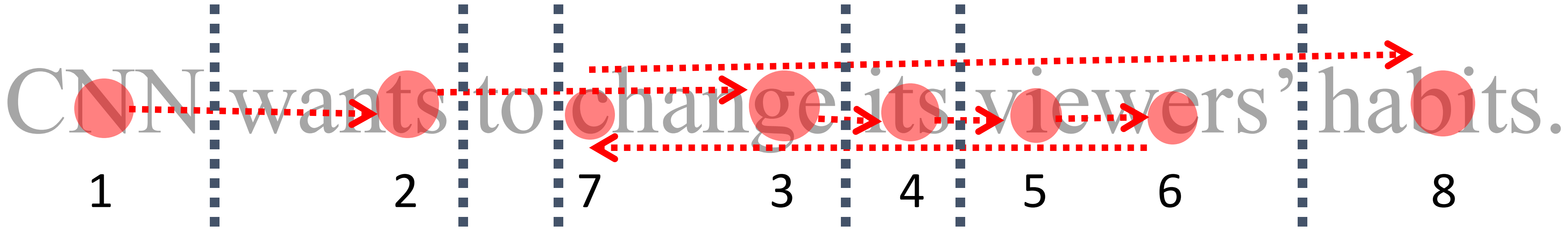
1. Extract fixations from raw samples

Eye Tracking: Preprocessing



1. Extract fixations from raw samples
2. Map fixations to *Areas of Interest:* Pre-defined screen areas (in pixels)

Eye Tracking: Preprocessing



Raw Data → Scanpaths → AoI-based Reading Measures

1. Extract fixations from raw samples
2. Map fixations to *Areas of Interest:* Pre-defined screen areas (in pixels)

Python package with preprocessing algorithms

pymovements.readthedocs.io

Eye Tracking Data Structure

Raw Data → Scanpaths → AoI-based Reading Measures

Raw Data

- Time series
- Each row contains one raw sample
- *N* depends on sampling frequency

Time (ms)	x (pixels)	y (pixels)
1	151	372
2	150	371
3	152	374
4	151	370
…	…	…

Scanpaths

- Discrete chronological sequence
- Each row contains one fixation

idx	x (mean, pixels)	y (mean, pixels)	word	aoi	dur
1	151	371	CNN	1	380
2	175	376	wants	2	180
3	198	378	change	4	224
4	227	370	viewers	6	299
5	251	369	habits	7	230
6	192	374	change	4	229
…	…	…	…	…	…

AoI-based Reading Measures

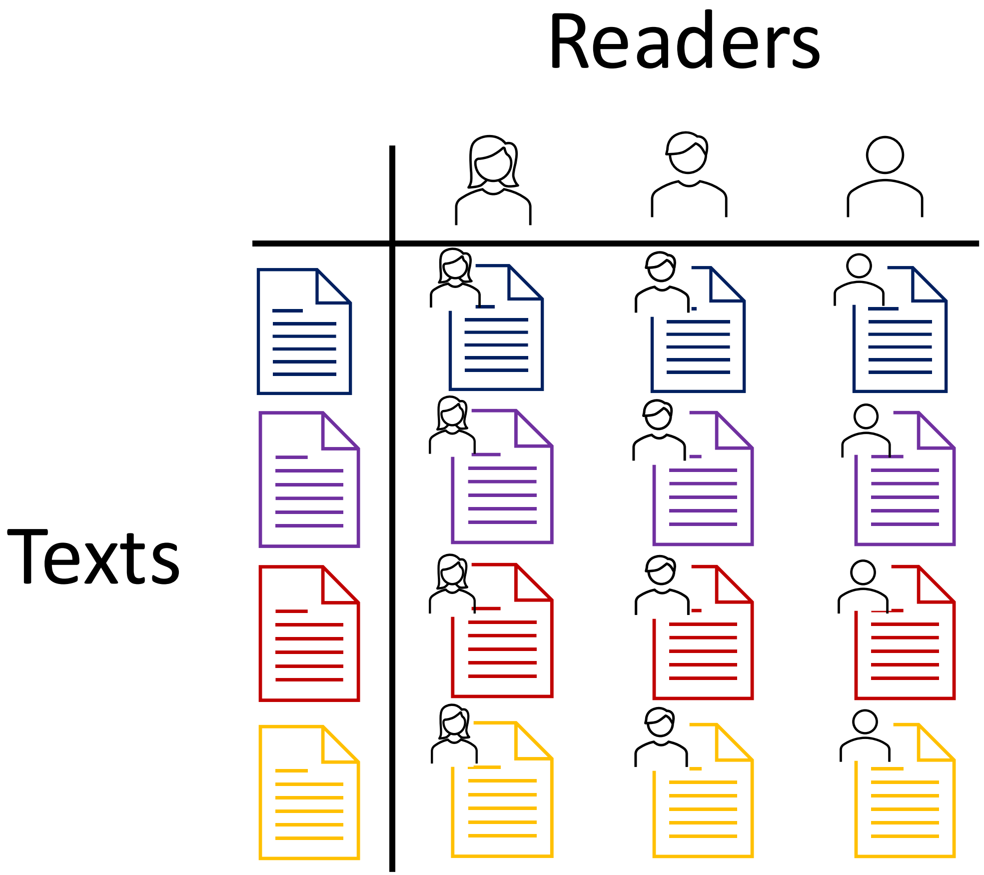
- Discrete sequence in aoi-order
- Each row contains RMs of one aoi read by one subj

subj	item	word	aoi	FFD
1	1	CNN	1	380
1	1	wants	2	180
1	1	to	3	NA
1	1	change	4	224
1	1	its	5	NA
1	1	viewers	6	299
1	1	habits	7	230
…	…	…	…	…

Data is **not** iid – it has **structure**

Implications for:

- Statistical modeling
- Training and evaluations
- Applications



Introduction to Eye Movements in Reading and Eye Tracking

How do People Read?

Data Representation

Reading Measures

Alternatives

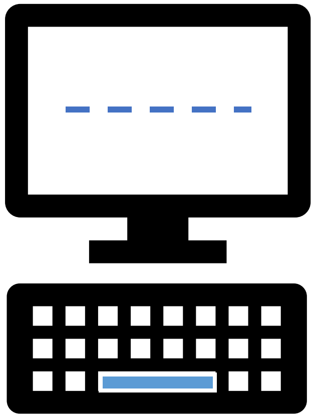
Eye Tracking

Datasets

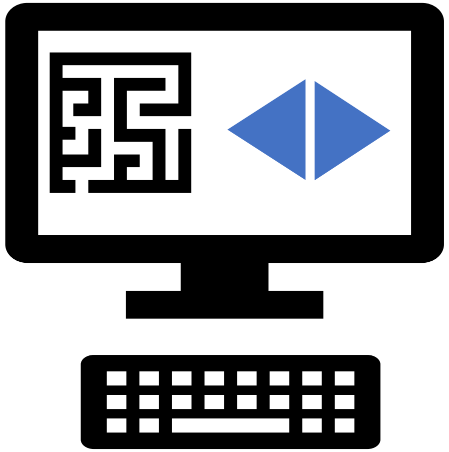
Eye Tracking vs Cheaper Low Tech Methods

Do we really need eye tracking?

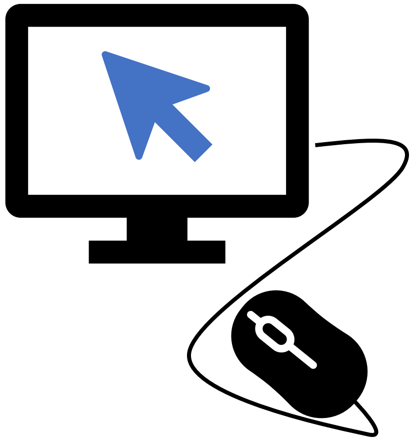
Popular alternatives in psycholinguistics:



Self-Paced Reading



Maze

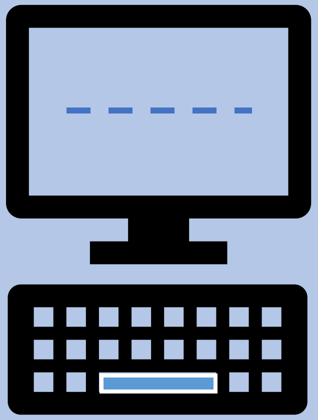


Mouse tracking

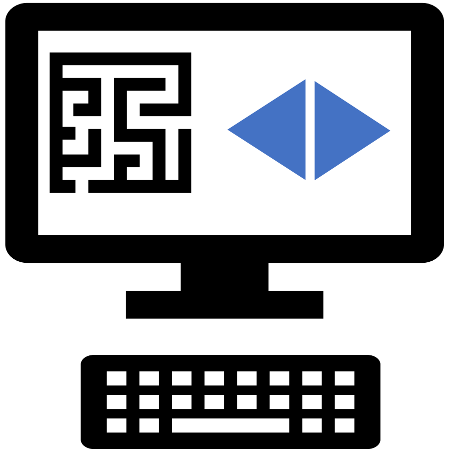
Eye Tracking vs Cheaper Low Tech Methods

Do we really need eye tracking?

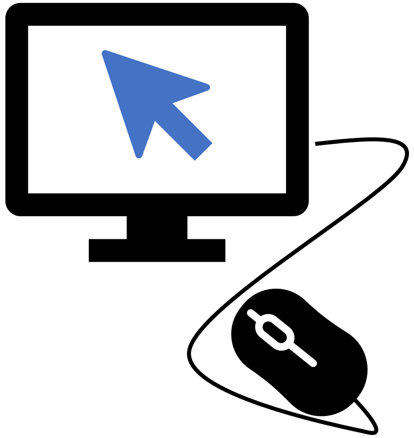
Popular alternatives in psycholinguistics:



Self-Paced Reading



Maze



Mouse tracking

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

_ _

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

```
Many--------------------------------------
```

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

-----years--------------------------

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

-----------later---------------------------

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

-----------------as-----------------------

Self Paced Reading (SPR)

- Reveal each consecutive word with a button press

---------------------he--------------------

Self Paced Reading (SPR)

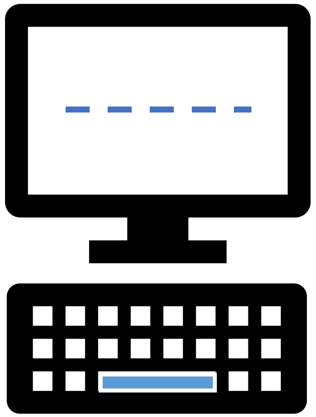
- Reveal each consecutive word with a button press
- Time between button presses as a proxy for incremental processing difficulty

------------------------faced--------------

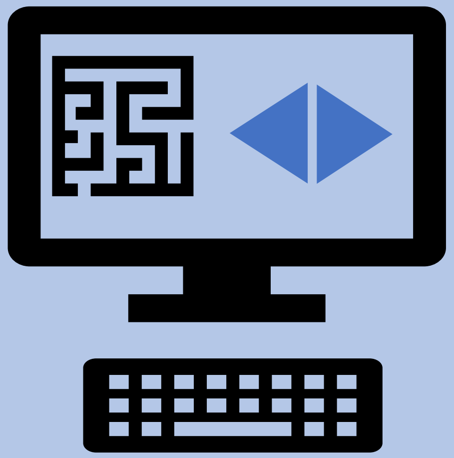
Eye Tracking vs Cheaper Low Tech Methods

Do we really need eye tracking?

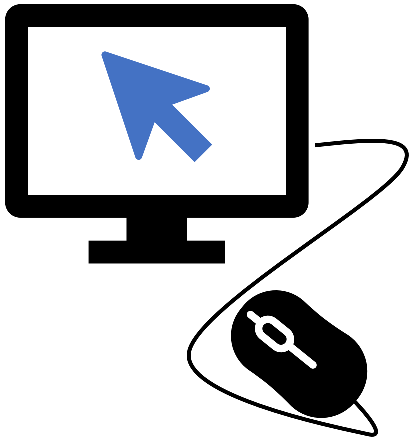
Popular alternatives in psycholinguistics:



Self Paced Reading



Maze



Mouse tracking

The Maze

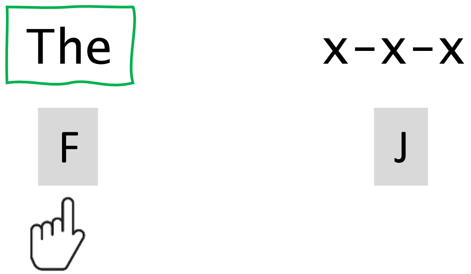
- Choose a word that fits given the preceding context

F J

[Forster et al. (2009)](), [Boyce et al. (2020)]()

The Maze

- Choose a word that fits given the preceding context



Forster et al. (2009), Boyce et al. (2020)

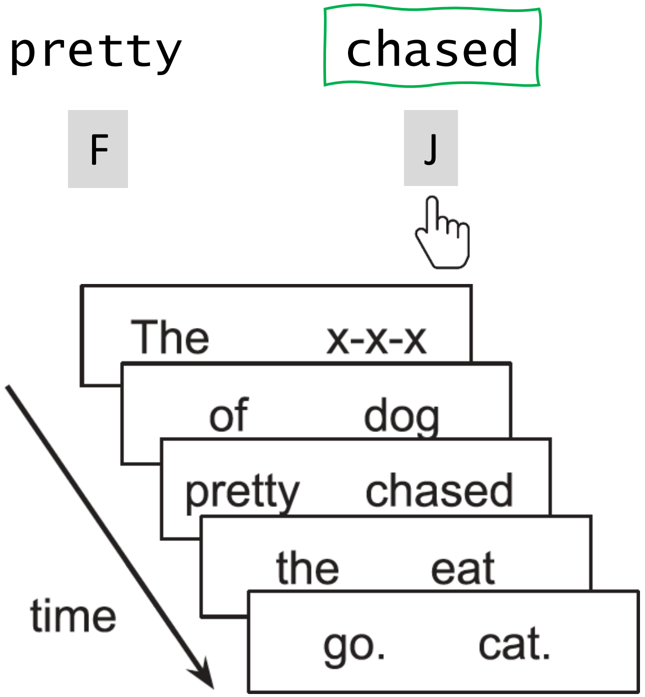
The Maze

- Choose a word that fits given the preceding context

Forster et al. (2009), Boyce et al. (2020)

The Maze

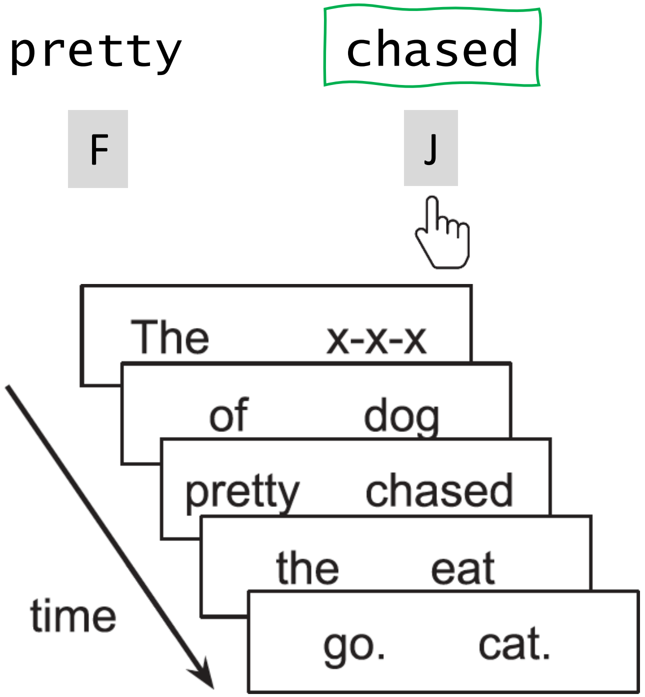
- Choose a word that fits given the preceding context



[Forster et al. (2009)](), [Boyce et al. (2020)]()

The Maze

- Choose a word that fits given the preceding context
- Time between button presses as a proxy for incremental processing difficulty

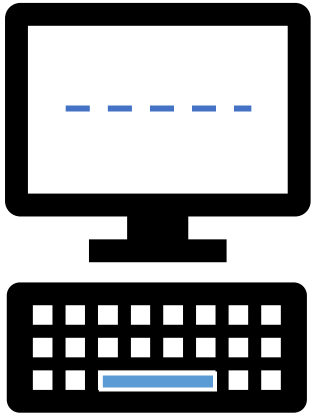


[Forster et al. (2009)](), [Boyce et al. (2020)]()

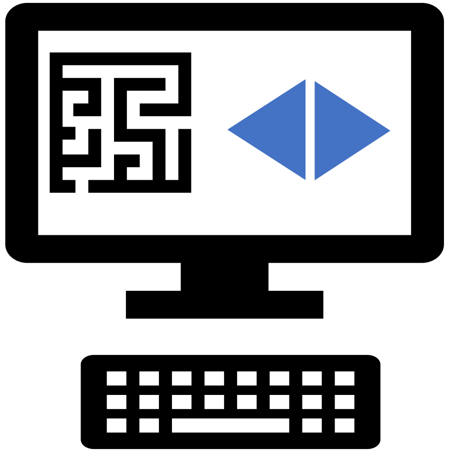
Eye Tracking vs Cheaper Low Tech Methods

Do we really need eye tracking?

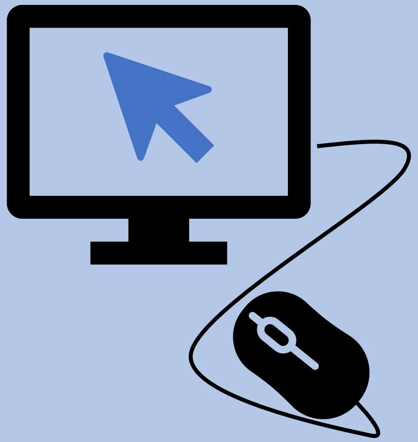
Popular alternatives in psycholinguistics:



Self Paced Reading

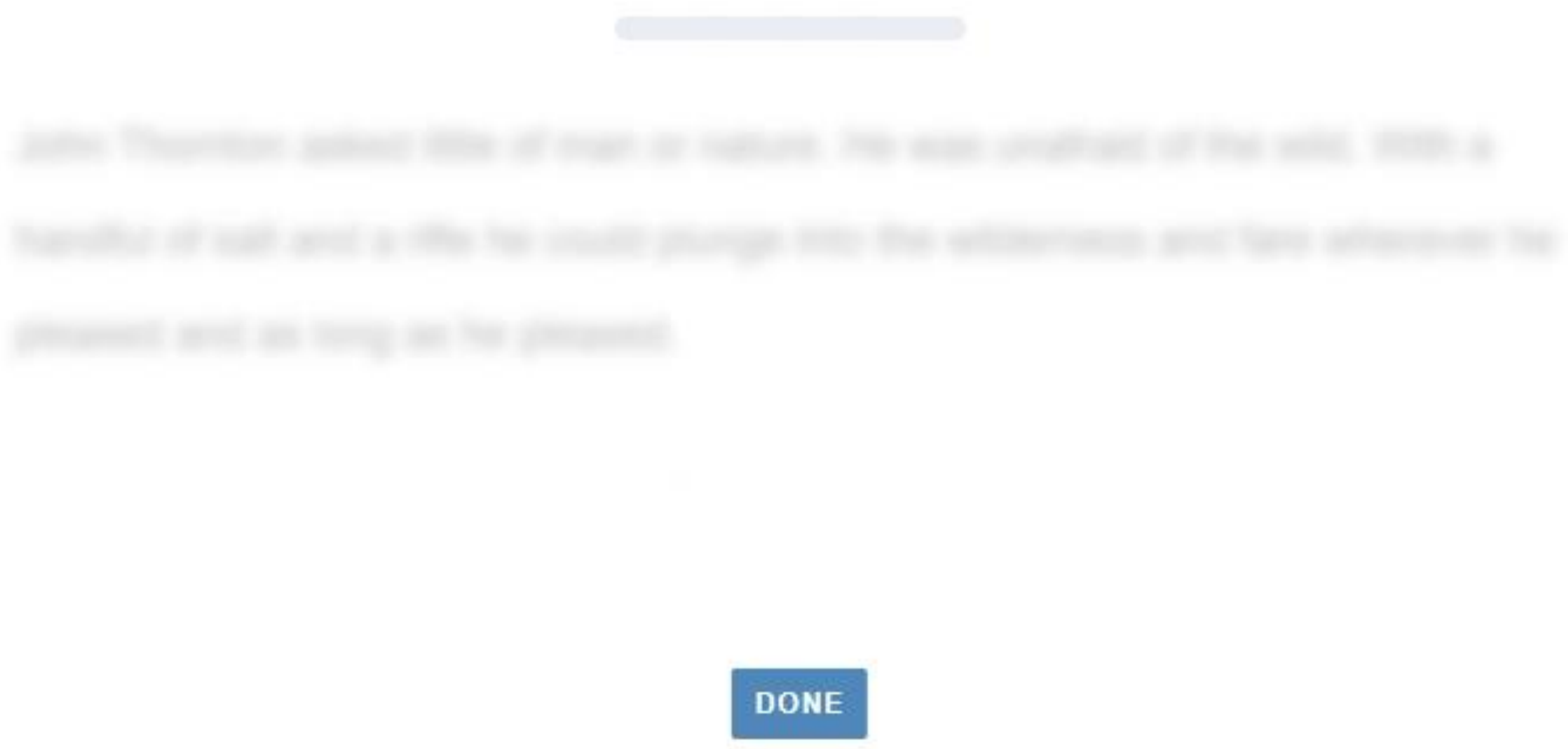


Maze



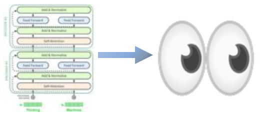
Mouse tracking

Mouse Tracking



Wilcox et al. (2024)

Synthetic Eye Tracking Data



Over the next 30 years, the planet's human population will increase to nine billion. Already one billion people do not get enough food. The increase will mean more pressure on agricultural land, water, forests, fisheries and biodiversity resources, as well as nutrients and energy supplies. There is also the issue of methane excreted by cows. The livestock farming contribution, in terms of greenhouse gas emissions, is enormous – 35% of the planet's methane, 65% of its nitrous oxide and 9% of the carbon dioxide.

1. **Cognitive models**
2. **NLP / ML Models**

Eye Tracking vs Cheaper Methods

- 🤩 More naturalistic
- 🤩 More fine-grained information (multiple measures, not just RTs)
- 🤩 Doesn’t include time to execute button presses and mouse movements
- 🤩 Higher quality than synthetic data
- 😟 Currently cannot be collected at scale (on the web)
- 😟 In most use cases, no eye tracking data is available at application time

Introduction to Eye Movements in Reading and Eye Tracking

How do People Read?

Data Representation

Reading Measures

Alternatives

Eye Tracking

Datasets

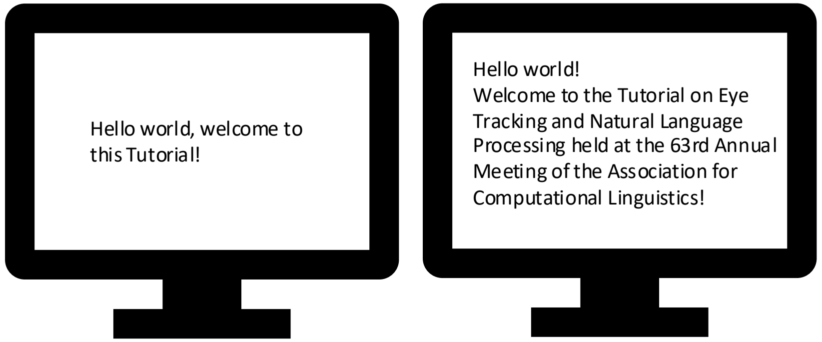
Types of Reading Datasets

Minimal-pairs vs naturalistic reading

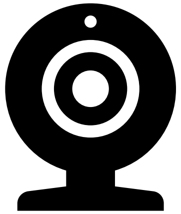
*1a) The horse raced past the barn **fell**.*

*1b) The horse **that was** raced past the barn **fell**.*

Single sentences vs parags/texts

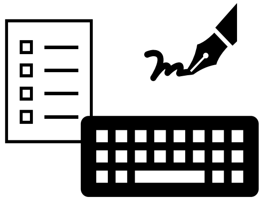


Eye tracking hardware quality



Reading task

- Natural reading
- Question answering
- Repeated reading
- ....

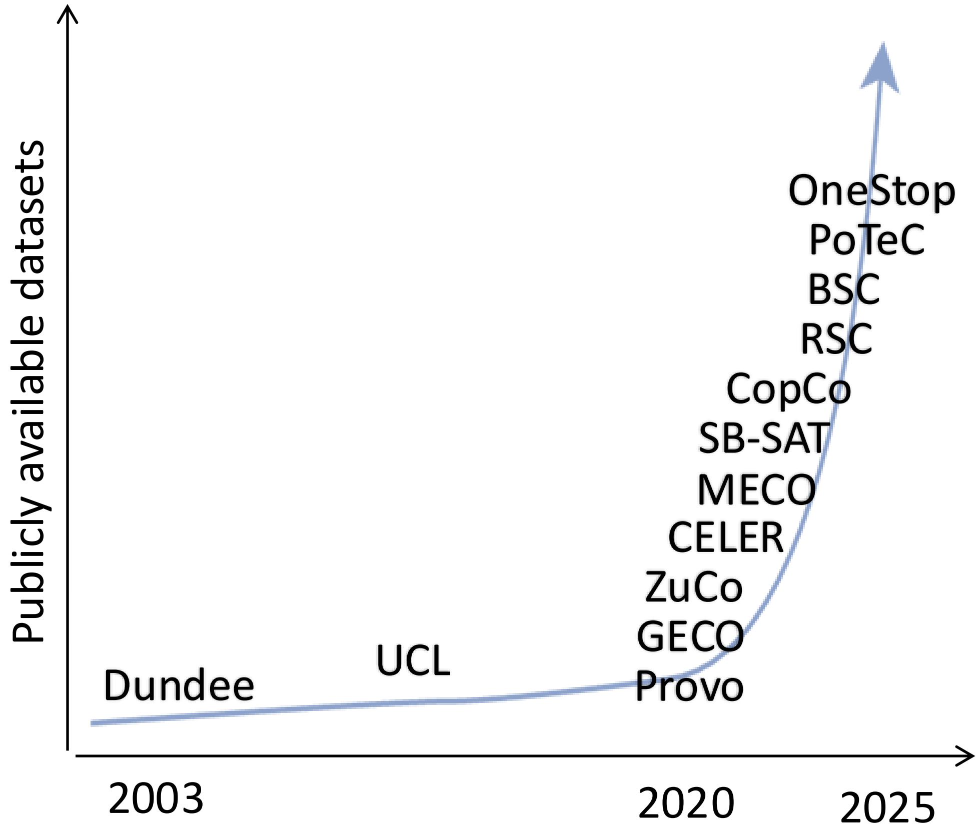


Additional Information Accompanying Eye Tracking Datasets

- Behavioral data
  - Response accuracies
  - Judgements/labels
- Psychometric test scores
- Demographic information
- Linguistic background

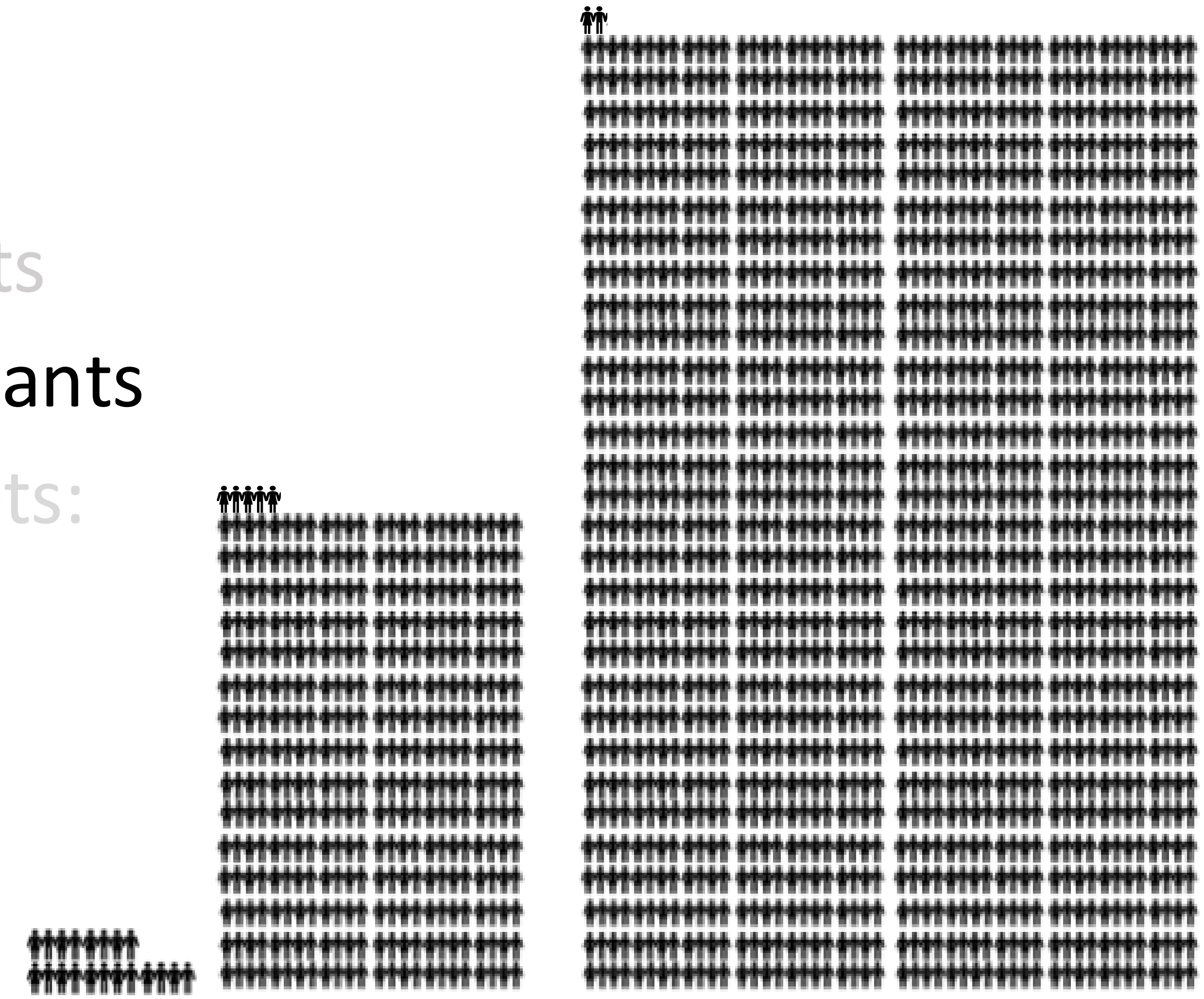
Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets



Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets



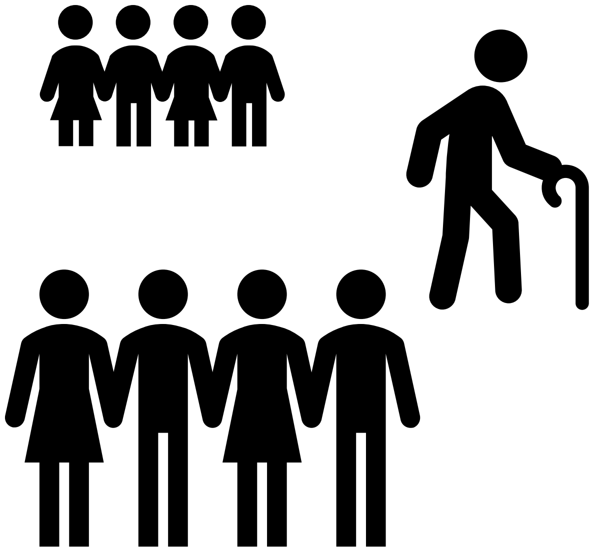
Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets

Chinese: BSC
Danish: CopCo
Dutch: GECO
German: PoTeC
Russian: RSC
...

Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets



Dyslexia: CopCo
L2: GECO, MECO, CELER, …
Wide age range: CELER
Domain-expertise: PoTeC

Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets

Reading comprehension: SB-SAT, OneStop, Multipleye
Sentiment classification: ETSA
Relation extraction: ZuCo
Repeated reading: OneStop
Information seeking: OneStop
...

Current Datasets

- Increasing number of data sets
- Increasing number of participants
- Increasing diversity of data sets:
  - Languages & Scripts
  - Populations
  - Tasks
- Multilingual data sets

MECO

MultiplEYE

Large-scale multi-lab multilingual data collection initiatives

MECO: https://meco-read.com

MultiplEYE: https://multipleye.eu

Overview of Publicly Available Data Sets

60+
Datasets

4.5K+
Participants Total

15K+
Text screens

30+
languages

How to Access the Datasets?

pymovements

Python package to download 24+ datasets with 4.5K+ participants

pymovements.readthedocs.io

Or from:

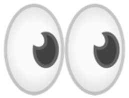
- Open science repositories
- Direct links from papers
- From authors' websites

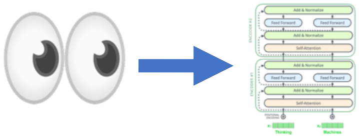
Datasets References

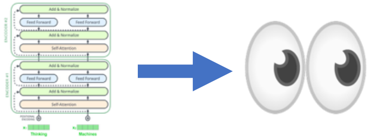
Introduction to Eye Movements in Reading and Eye Tracking

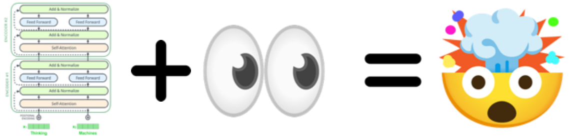
- How do People Read?
- Reading Measures
- Eye Tracking
- Data Representation
- Alternatives
- Datasets

Tutorial Outline


1. Introduction to eye tracking

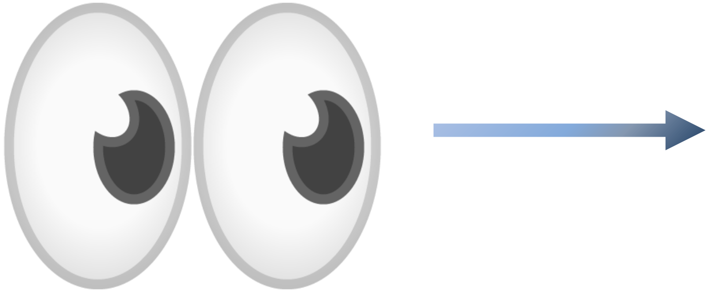

2. Uses of Eye Tracking in NLP

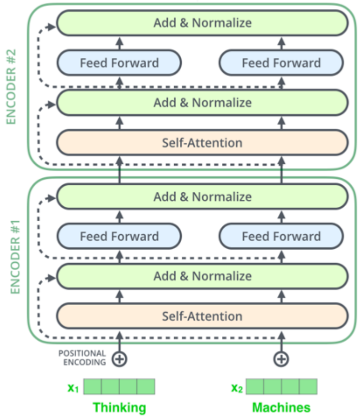

3. NLP for eye movement and cognitive modeling


4. New human centered applications


5. Outlook and future directions

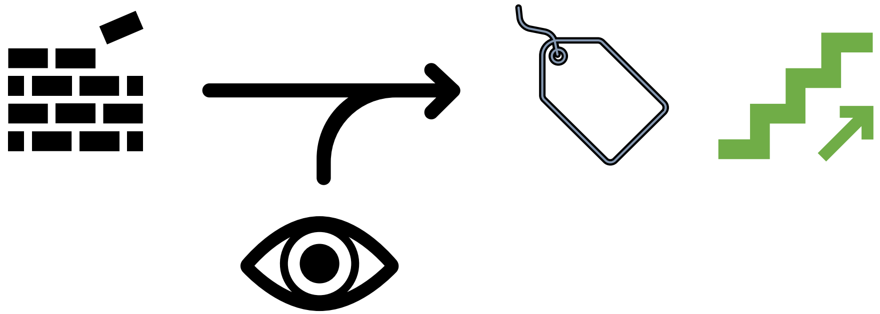
Uses of Eye Tracking in NLP





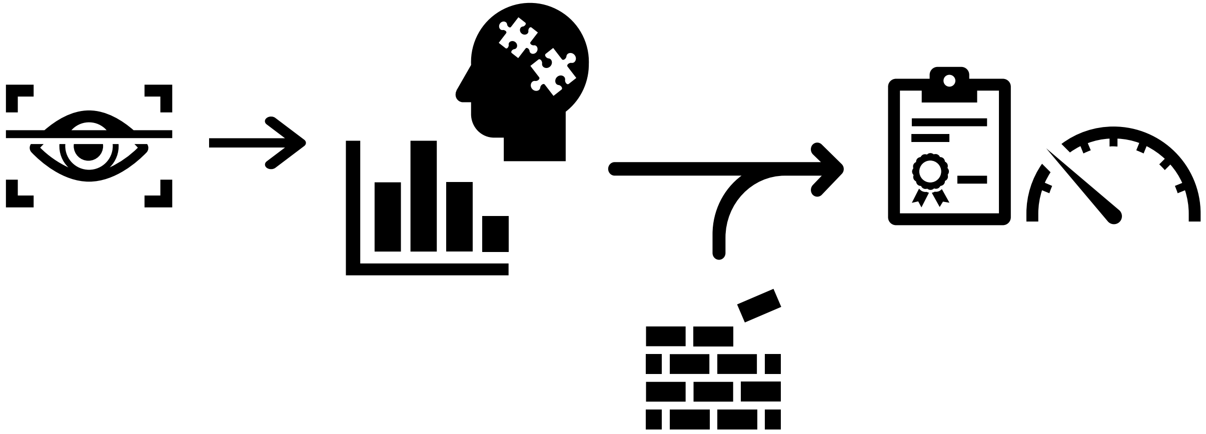
Uses of Eye Tracking in NLP 👀

Modeling



Eye movements can enhance the performance of NLP models

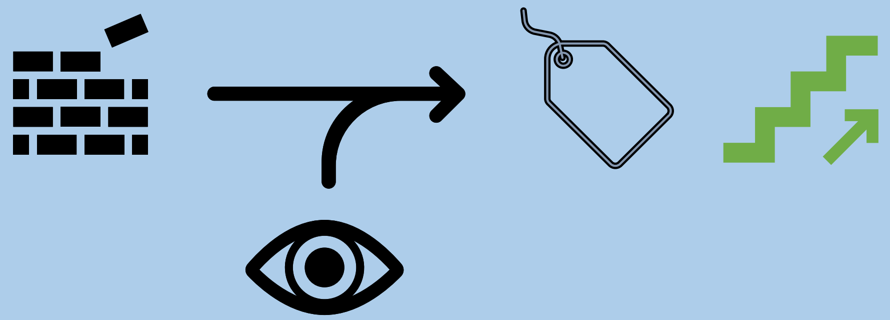
Evaluation



Eye movements as behavioral benchmarks for evaluating NLP models

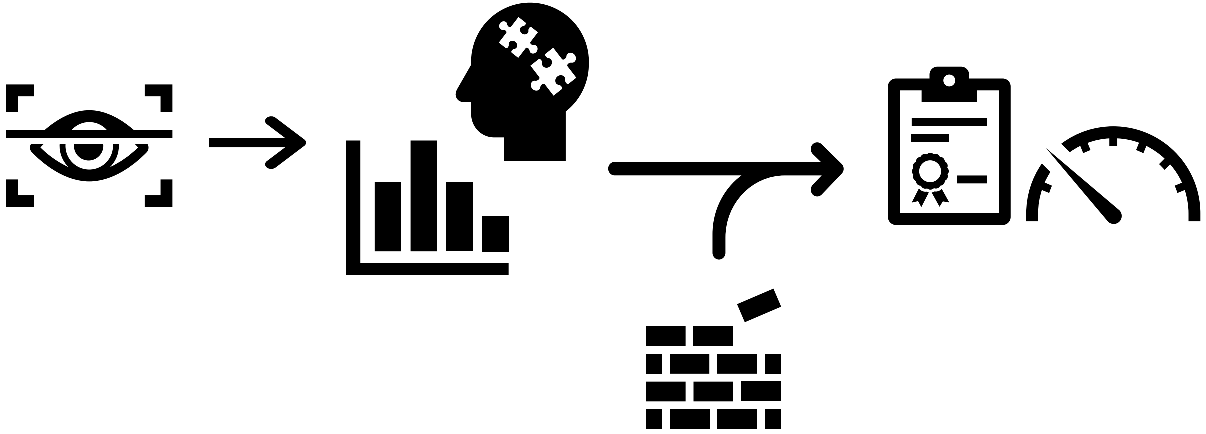
Uses of Eye Tracking in NLP

Modeling



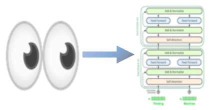
Eye movements can enhance the performance of NLP models

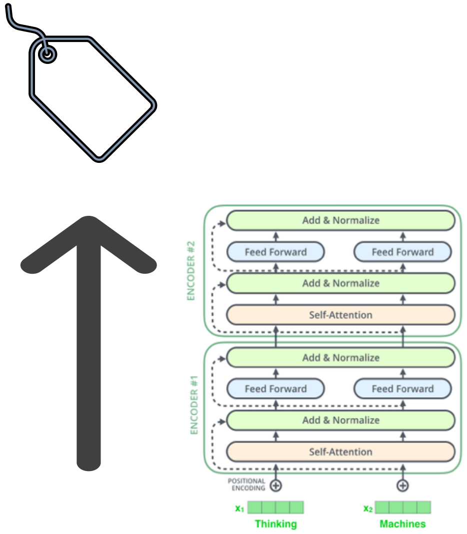
Evaluation



Eye movements as behavioral benchmarks for evaluating NLP models

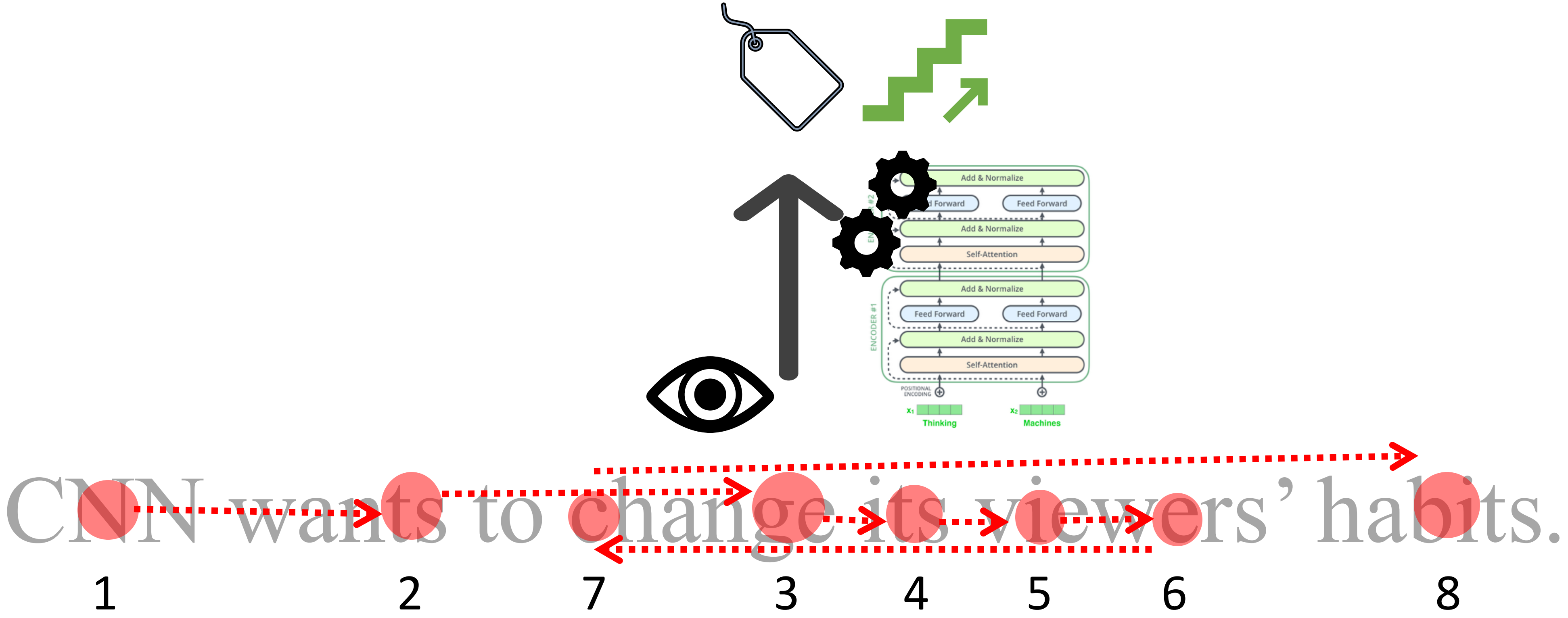
Language modeling





CNN wants to change its viewers’ habits.

Language modeling and eye movements 👀



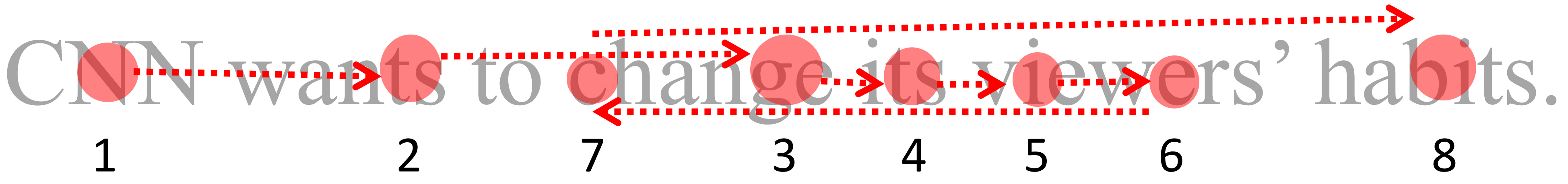
Improving NLP with Gaze

Tasks

- Sentiment analysis Mishra et al. (2016), Barrett et al. (2018), Yang and Hollenstein (2023)
- NER Hollenstein and Zhang (2019)
- Paraphrase generation, sentence compression Sood et al. (2020), Klerke et al. (2016)
- Relation extraction, sentiment analysis, NER Ren and Xiong (2021)
- GLUE Deng et al. (2023), Deng et al. (2024)
- Readability assessment González-Garduño and Søgaard (2017)
- Dependency parsing Strzyz et al. (2019)
- QA Malmaud et al. (2020)

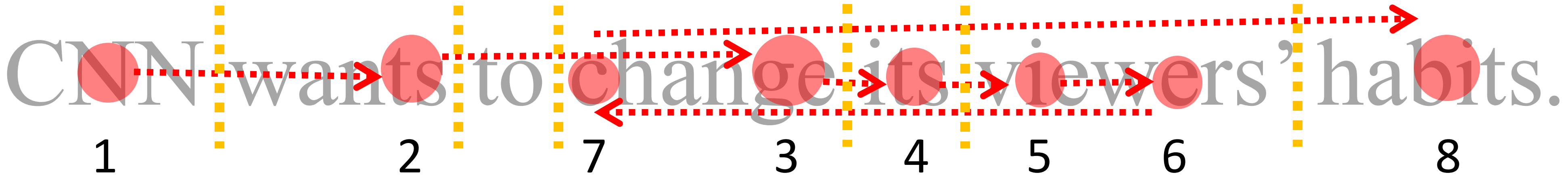
Why eye movements?

- Eye movement in reading are influenced by
  - the **difficulty** of the text
  - the **individual**
  - **cognitive demands**
  - …



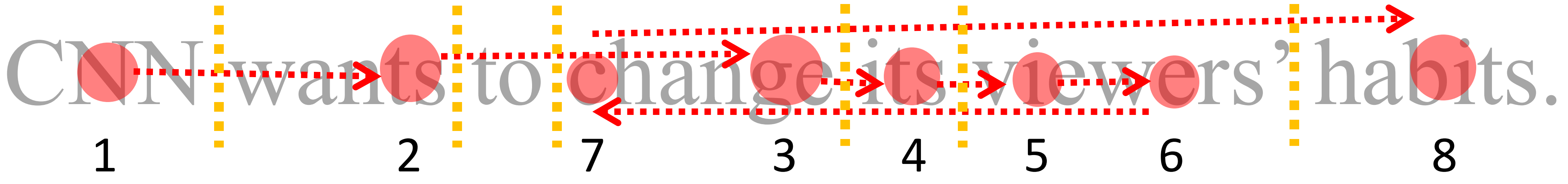
How to combine eye movements and text?

- Static text: linguistic axis
- Dynamic eye movements: temporal axis



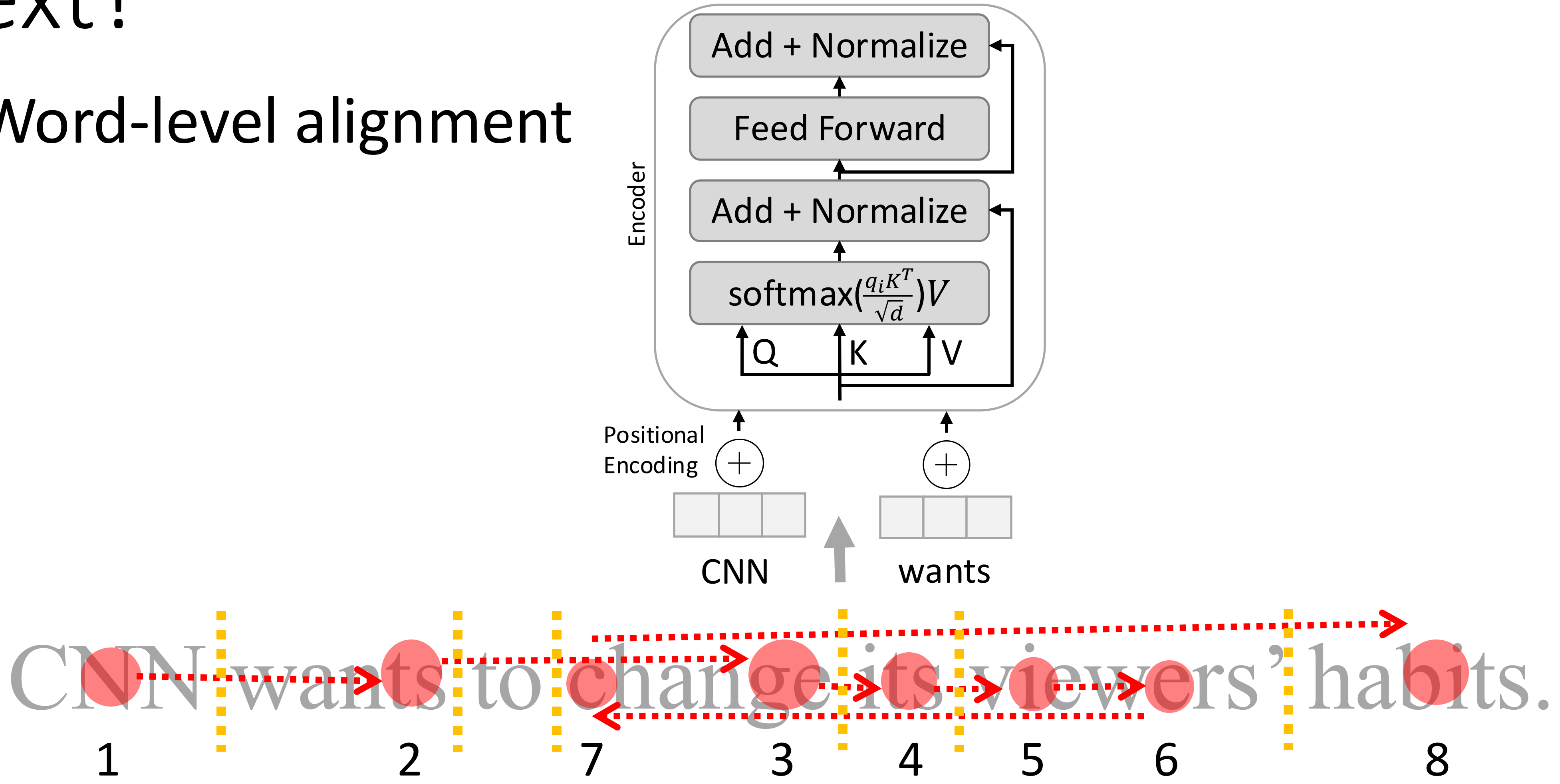
How to combine eye movements and text?

- Word-level alignment
  - E.g. total fixation duration



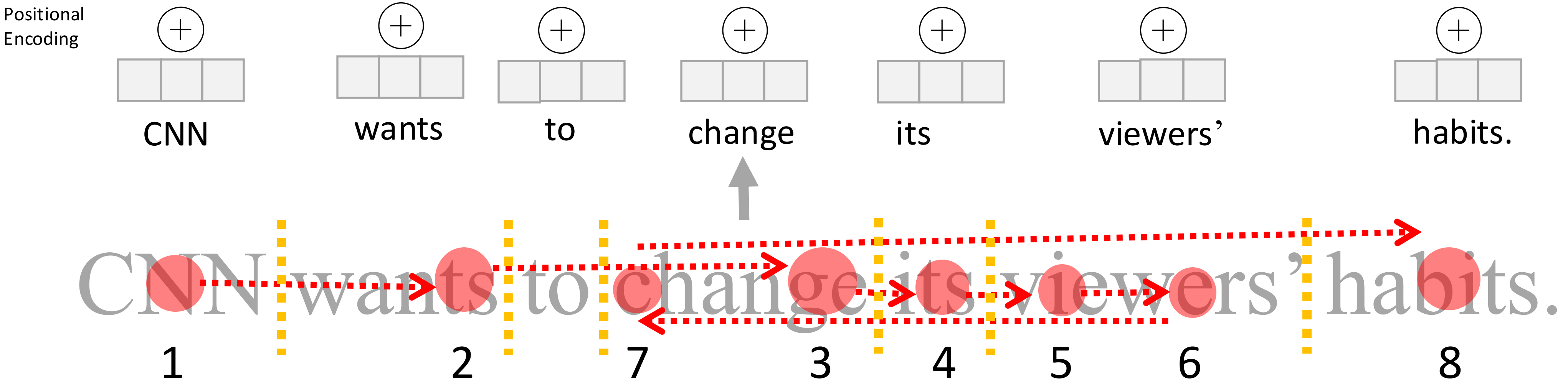
How to combine eye movements and text?

- Word-level alignment



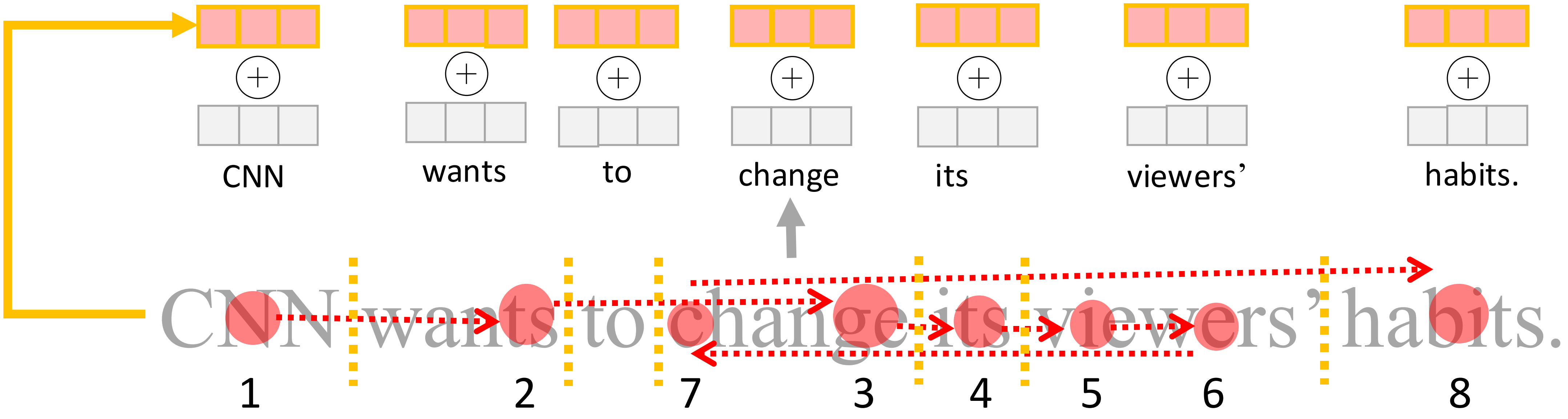
How to combine eye movements and text?

- Word-level alignment



How to combine eye movements and text?

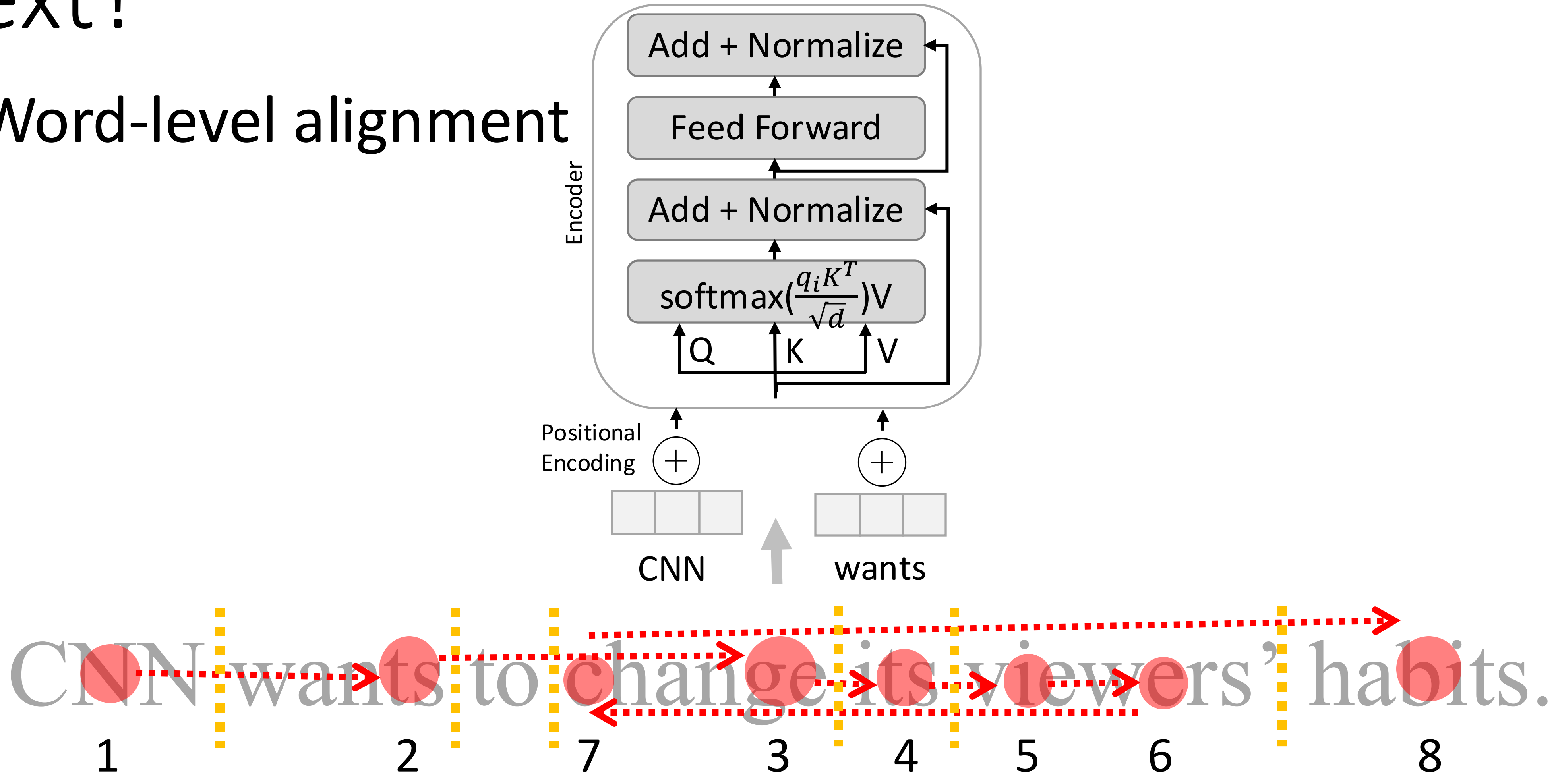
- Word-level eye movement embedding



López-Cardona et al. (2025)

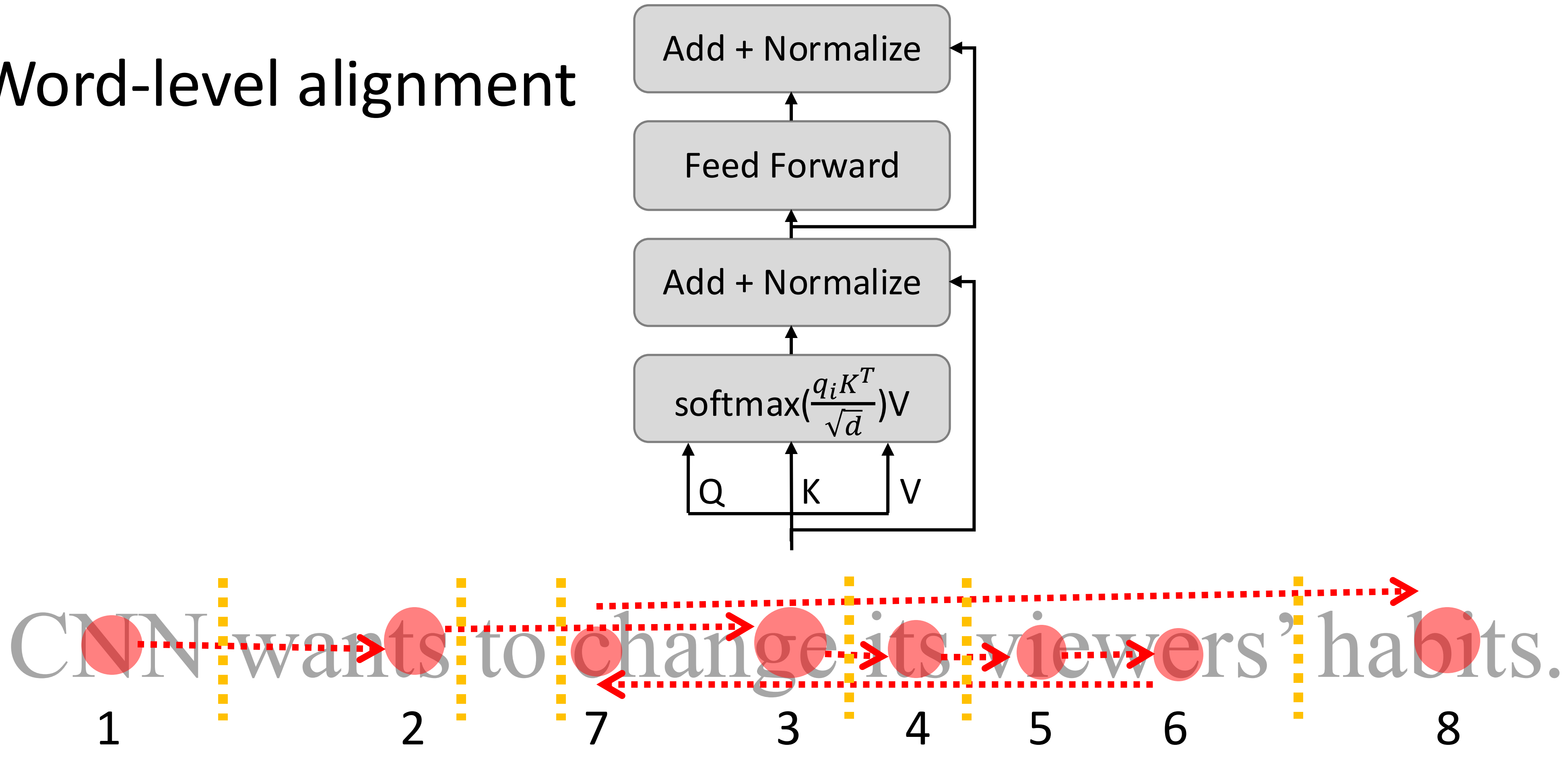
How to combine eye movements and text?

- Word-level alignment



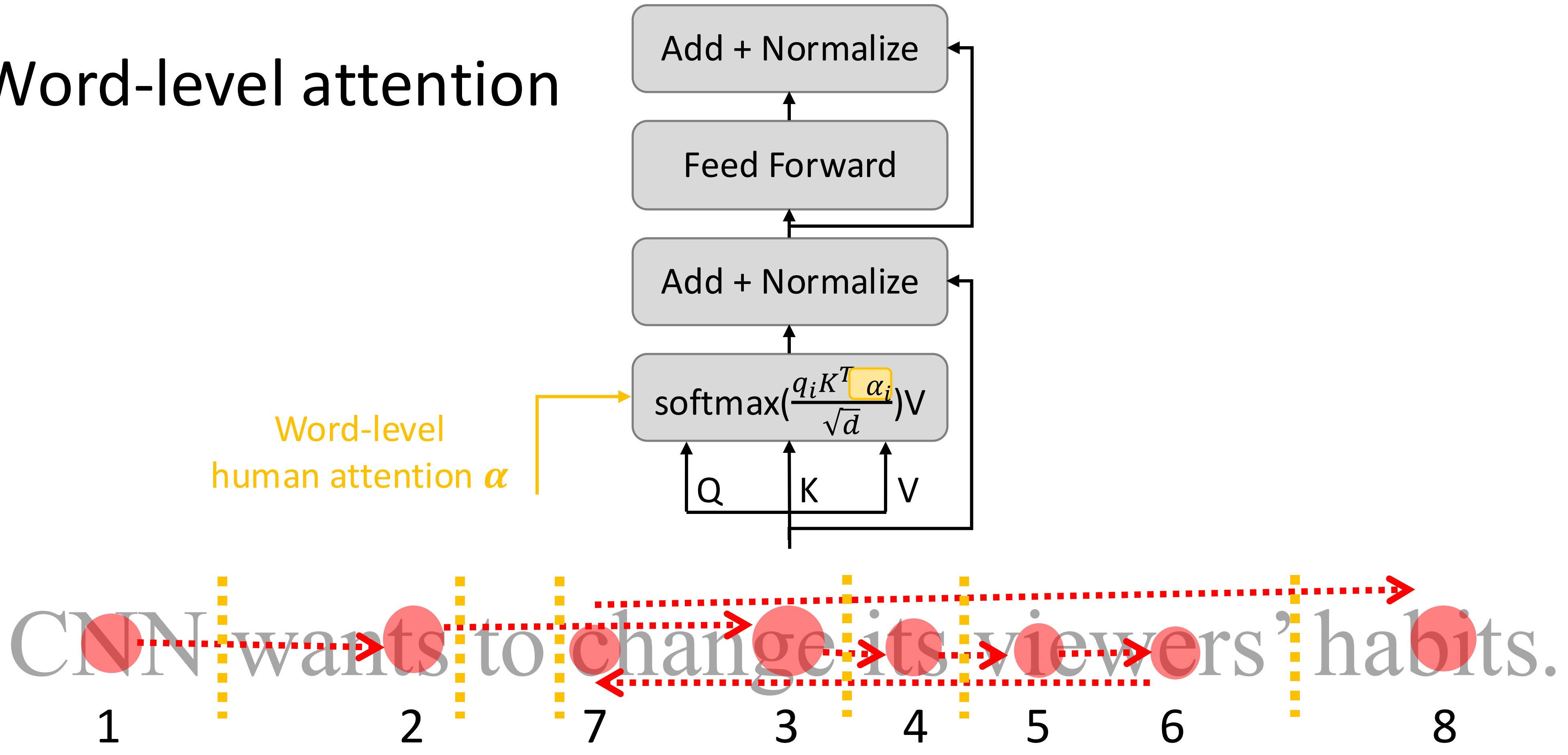
How to combine eye movements and text?

- Word-level alignment



How to combine eye movements and text?

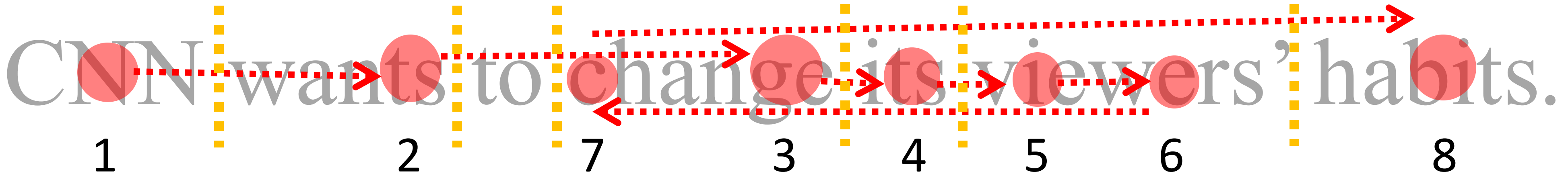
- Word-level attention



Sood et al. (2022)

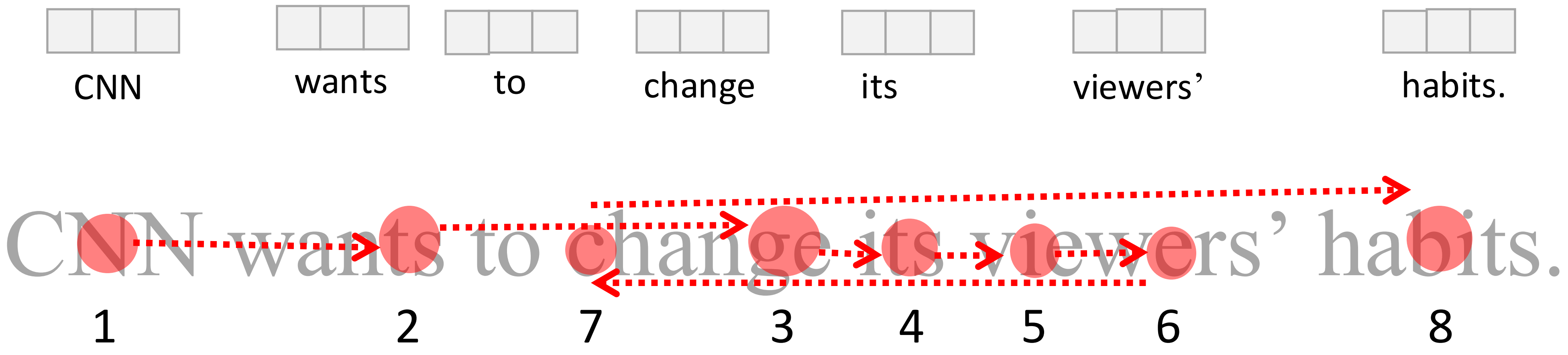
How to combine eye movements and text?

- Word-level alignment
  - E.g. total fixation duration



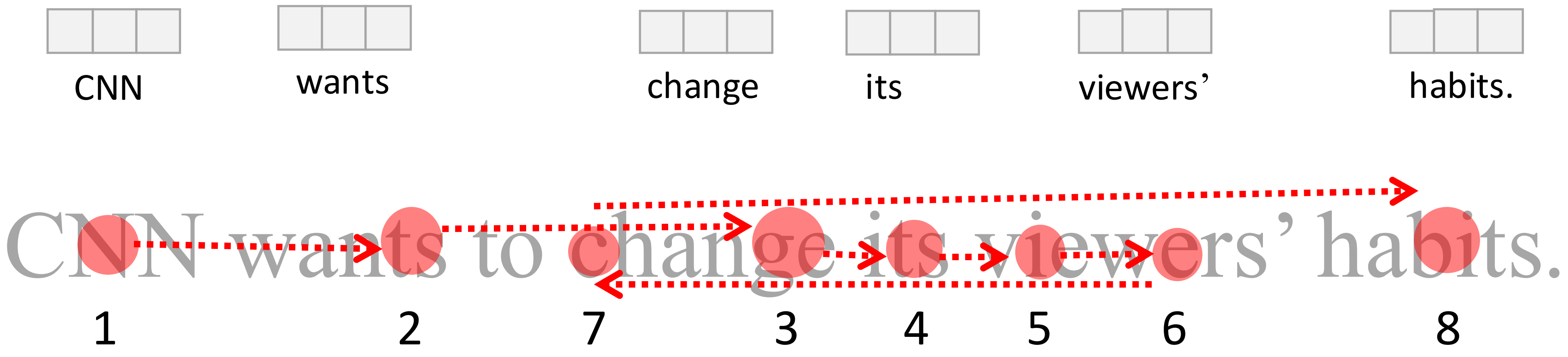
How to combine eye movements and text?

- Word reordering



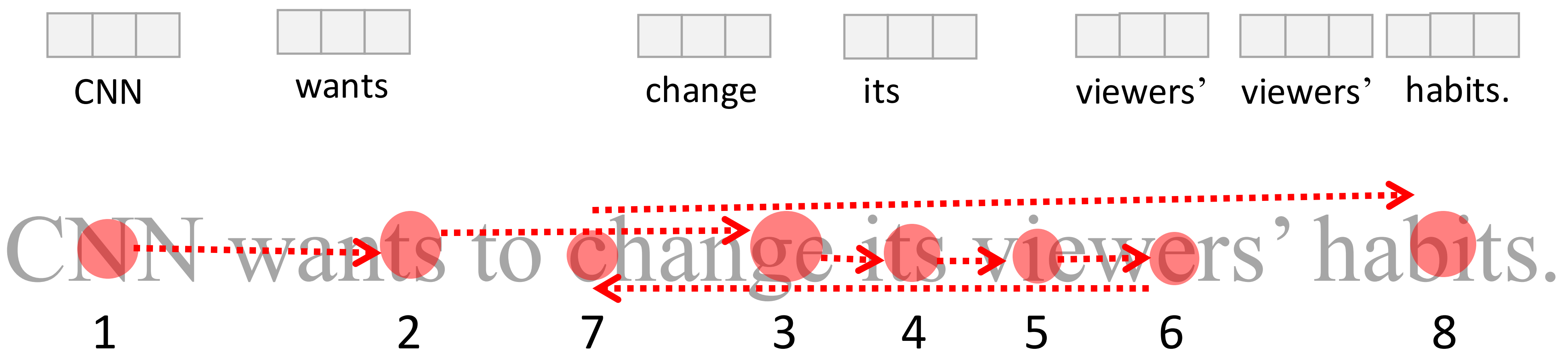
How to combine eye movements and text?

- Word reordering



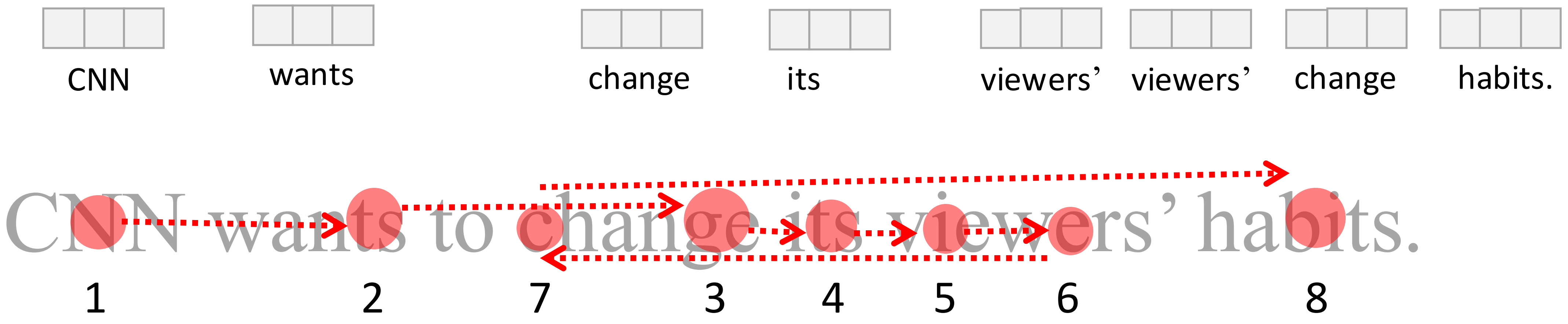
How to combine eye movements and text?

- Word reordering



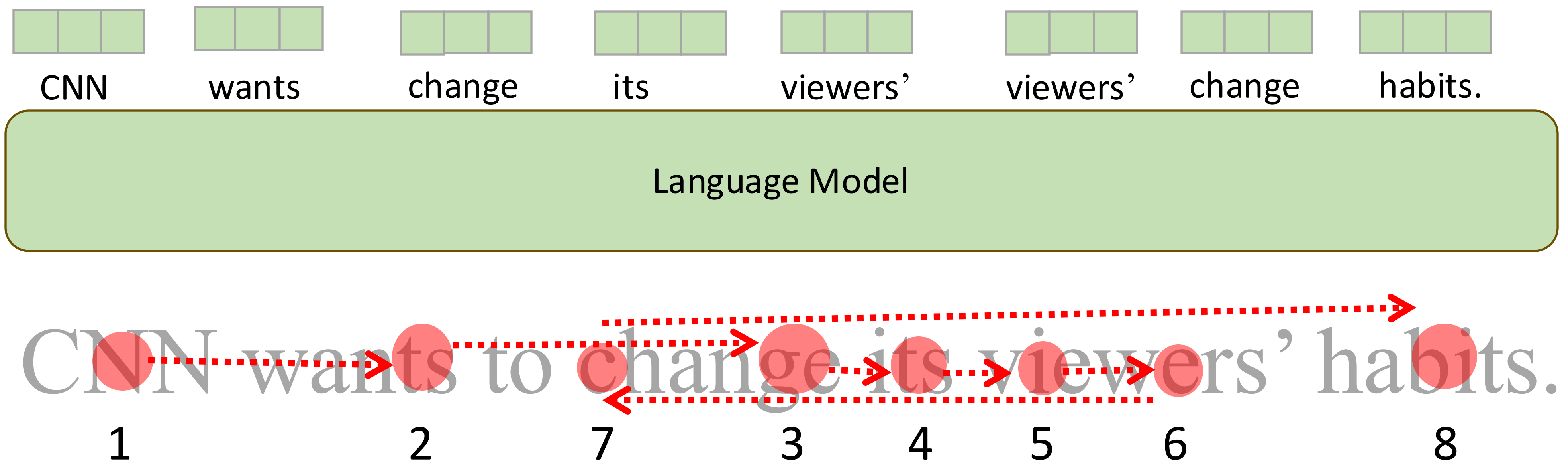
How to combine eye movements and text?

- Word reordering



How to combine eye movements and text?

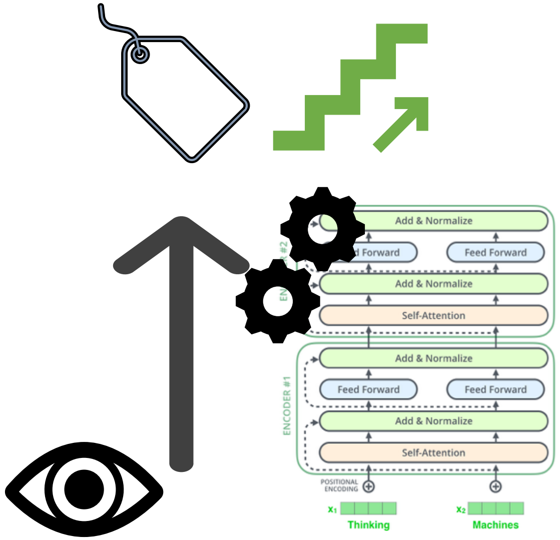
- Word reordering

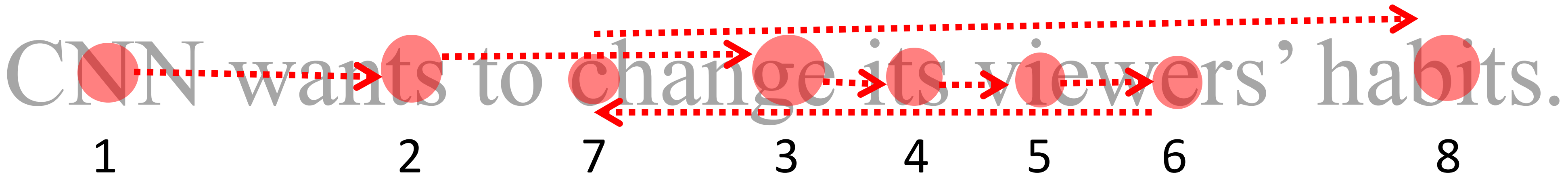


Yang and Hollenstein (2023), Deng et al. (2023), Deng et al. (2024)

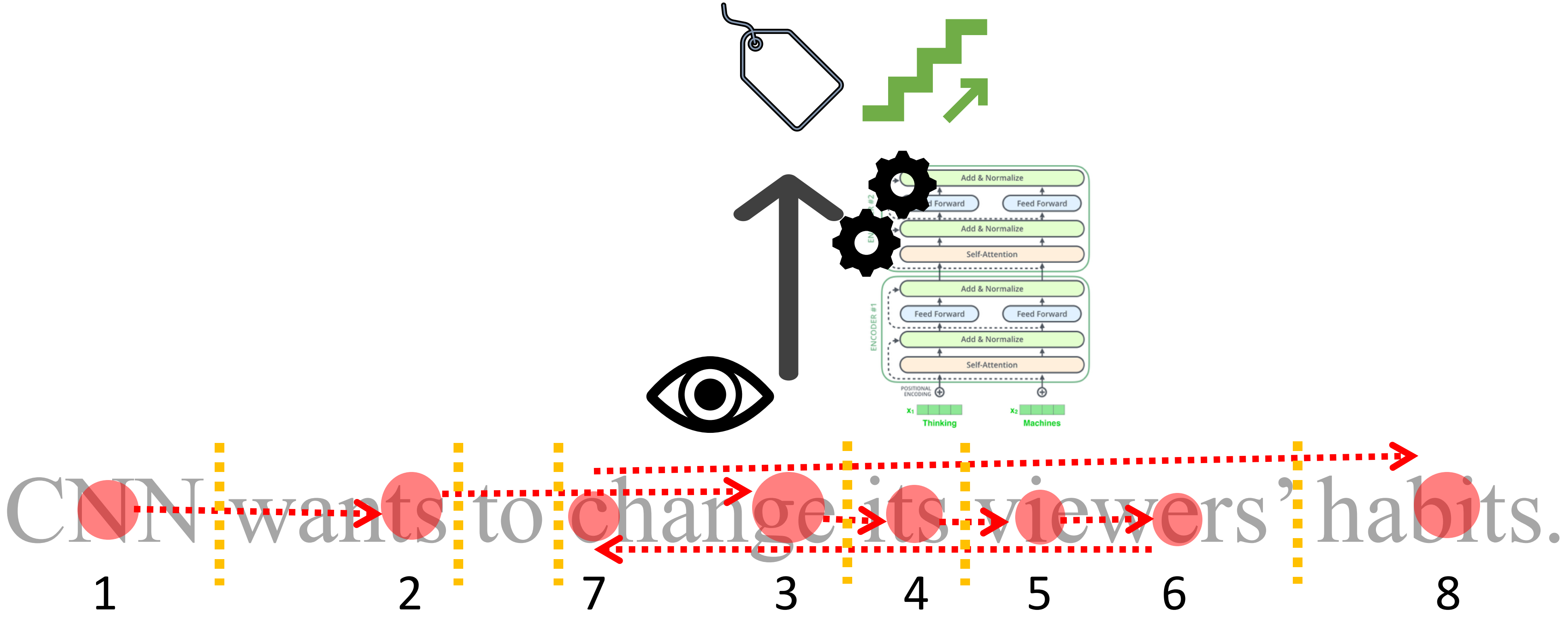
How to combine eye movements and text?

- Input:
  - Positional encoding
  - Attention
  - Reordering



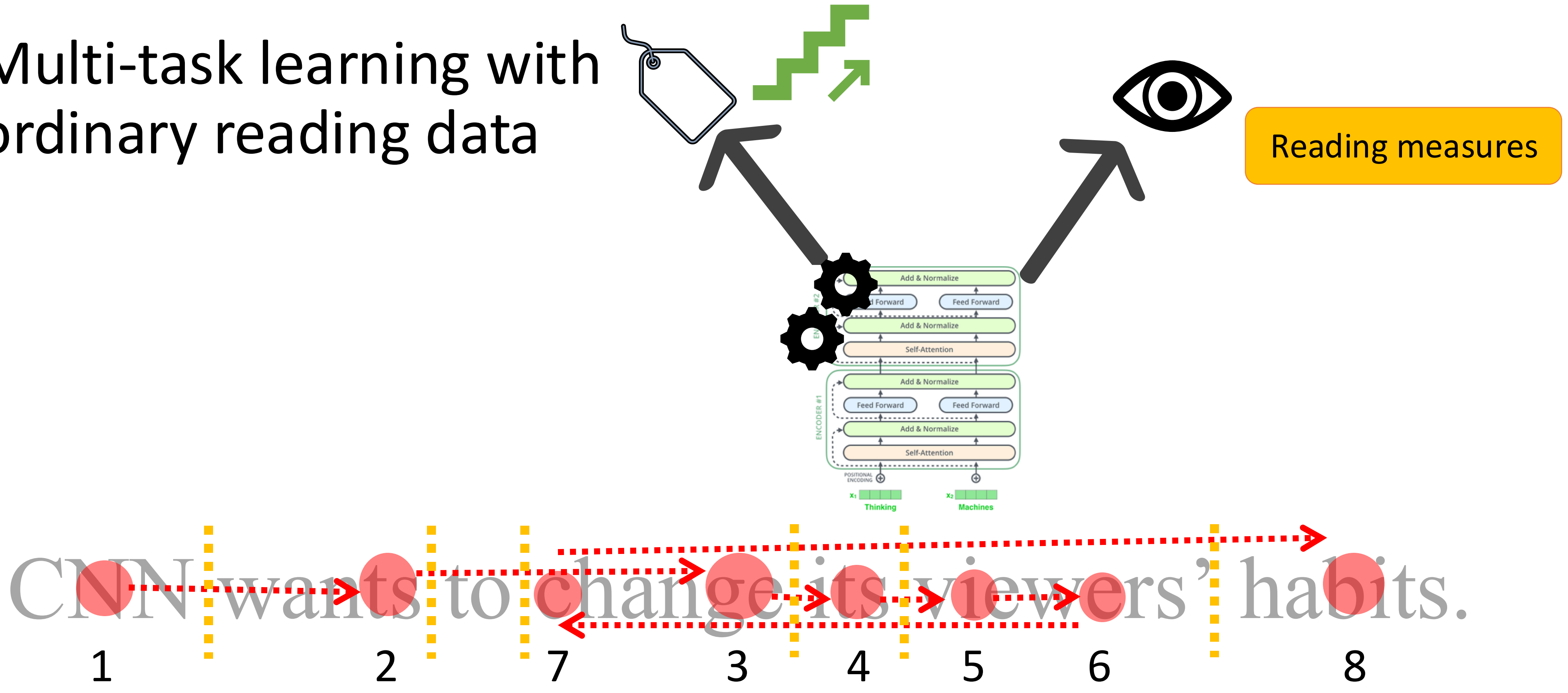


How to combine eye movements and text?



How to combine eye movements and text?

- Multi-task learning with ordinary reading data



Klerke et al. (2016), González-Garduño and Søgaard (2017), Barrett et al. (2018), Strzyz et al. (2019)

How to combine eye movements and text?

- Multi-task learning with task specific (QA) data Malmaud et al. (2020)

Ordinary reading (no question preview)

In	the	next	30	years,	the	planet's	human	population	will
increase	to	nine	billion.	Already	one	billion	people	do	not
get	enough	food.	The	increase	will	put	more	pressure	on
agricultural	land,	water,	forests,	fisheries	and	resources,	and	also	food
and	energy	supplies.	The	cost	of	meat	is	increasing	–
it	costs	more	money	now,	but	also	people	have	to
destroy	a	lot	of	rainforest	to	make	fields	or	to
grow	food	for	cows.	Cows	also	make	methane.	The	farming
of	cows,	pigs	and	sheep	makes	very	large	amounts	of
greenhouse	gases	–	35%	of	the	planet's	methane,	65%	of
its	nitrous	oxide	and	9%	of	the	carbon	dioxide.	

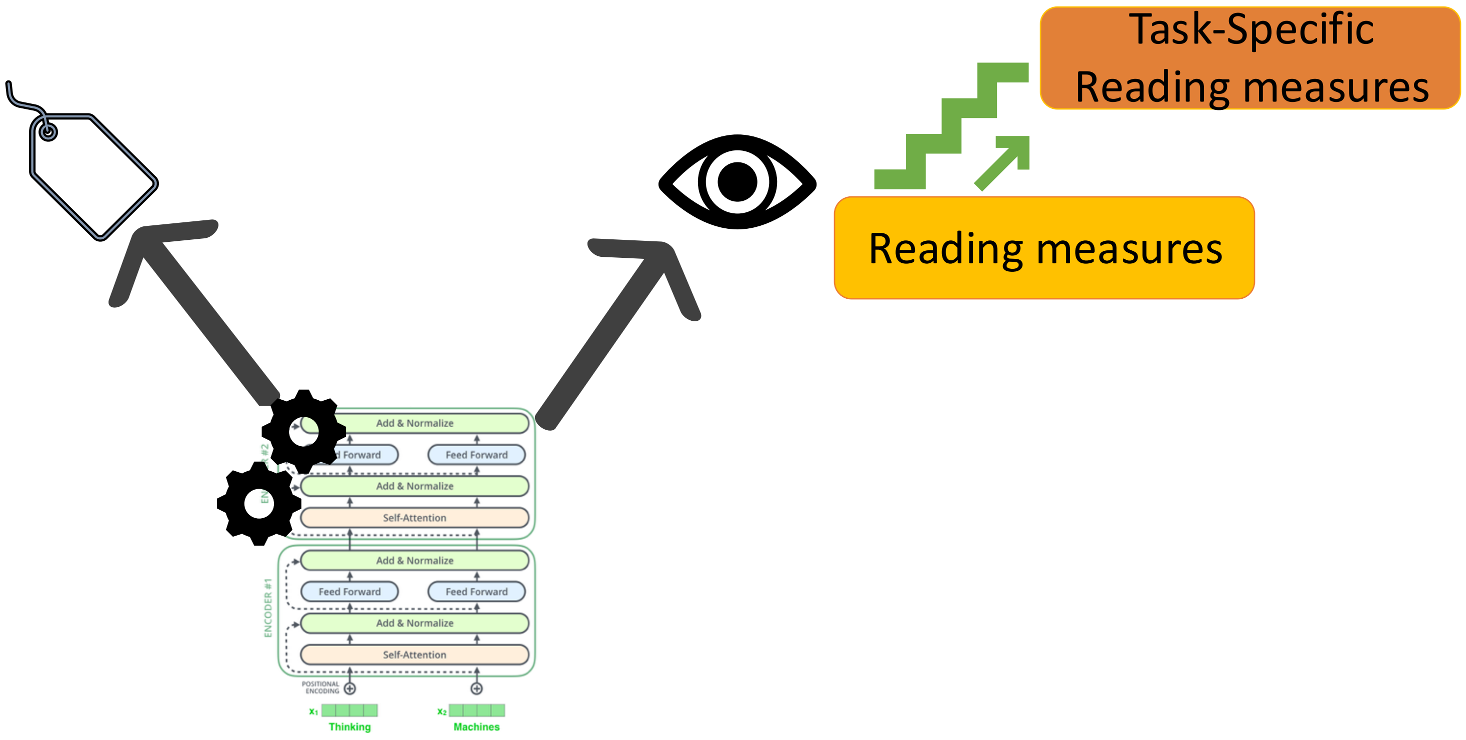
Information seeking (with question preview)

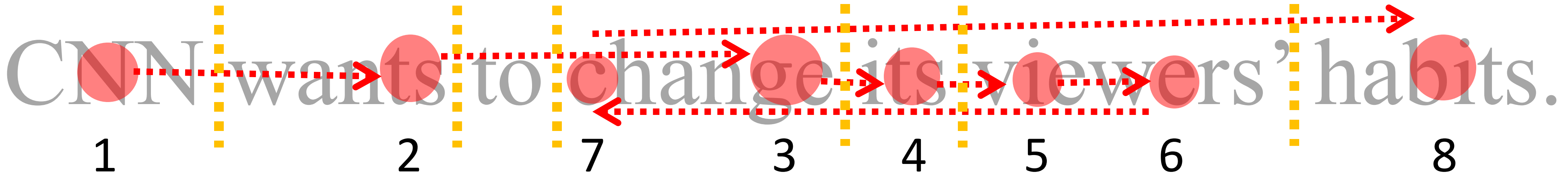
Q: What will result from an increase in human population in the future?

In	the	next	30	years,	the	planet's	human	population	will
increase	to	nine	billion.	Already	one	billion	people	do	not
get	enough	food.	The	increase	will	put	more	pressure	on
agricultural	land,	water,	forests,	fisheries	and	resources,	and	also	food
and	energy	supplies.	The	cost	of	meat	is	increasing	–
it	costs	more	money	now,	but	also	people	have	to
destroy	a	lot	of	rainforest	to	make	fields	or	to
grow	food	for	cows.	Cows	also	make	methane.	The	farming
of	cows,	pigs	and	sheep	makes	very	large	amounts	of
greenhouse	gases	–	35%	of	the	planet's	methane,	65%	of
its	nitrous	oxide	and	9%	of	the	carbon	dioxide.	

How to combine eye movements and text?

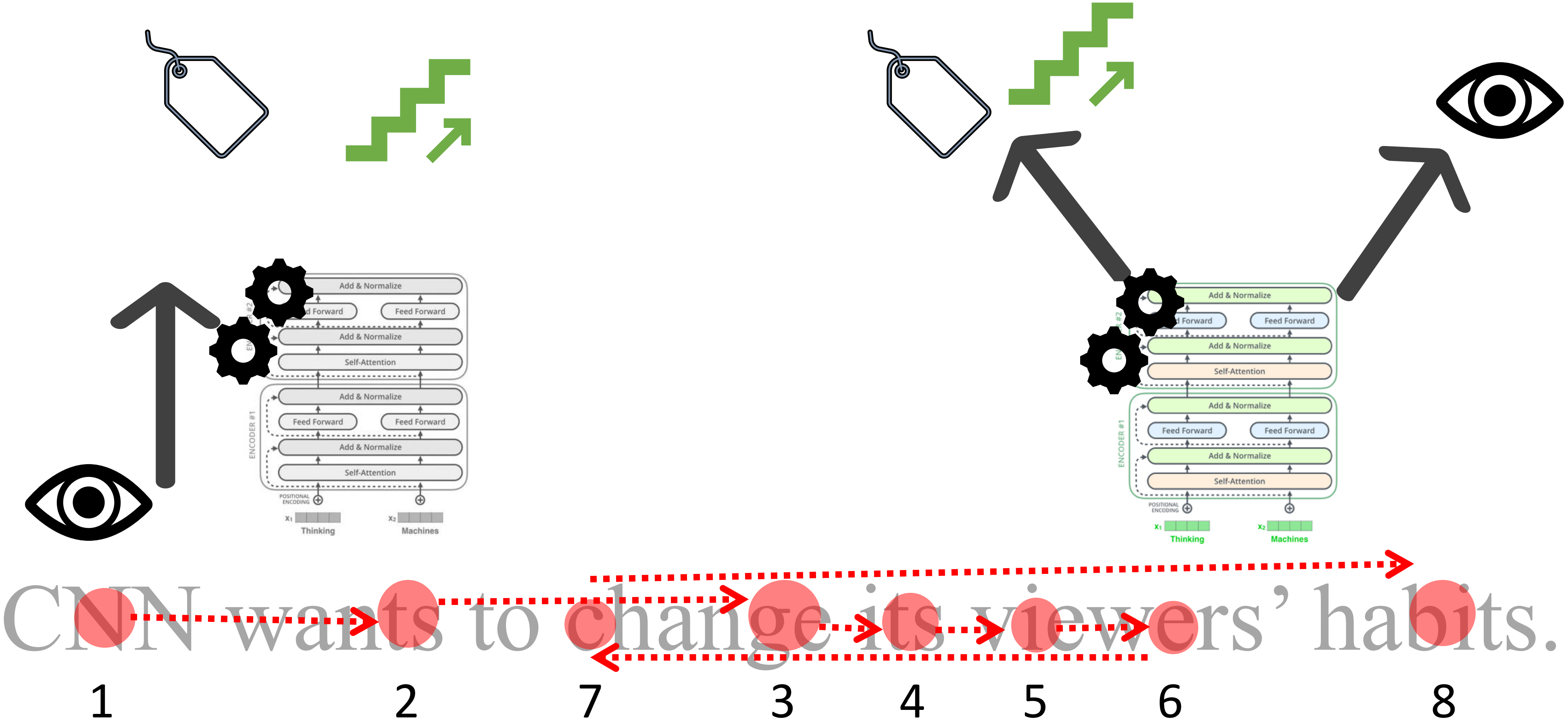
- Multi-task learning with task-specific reading data





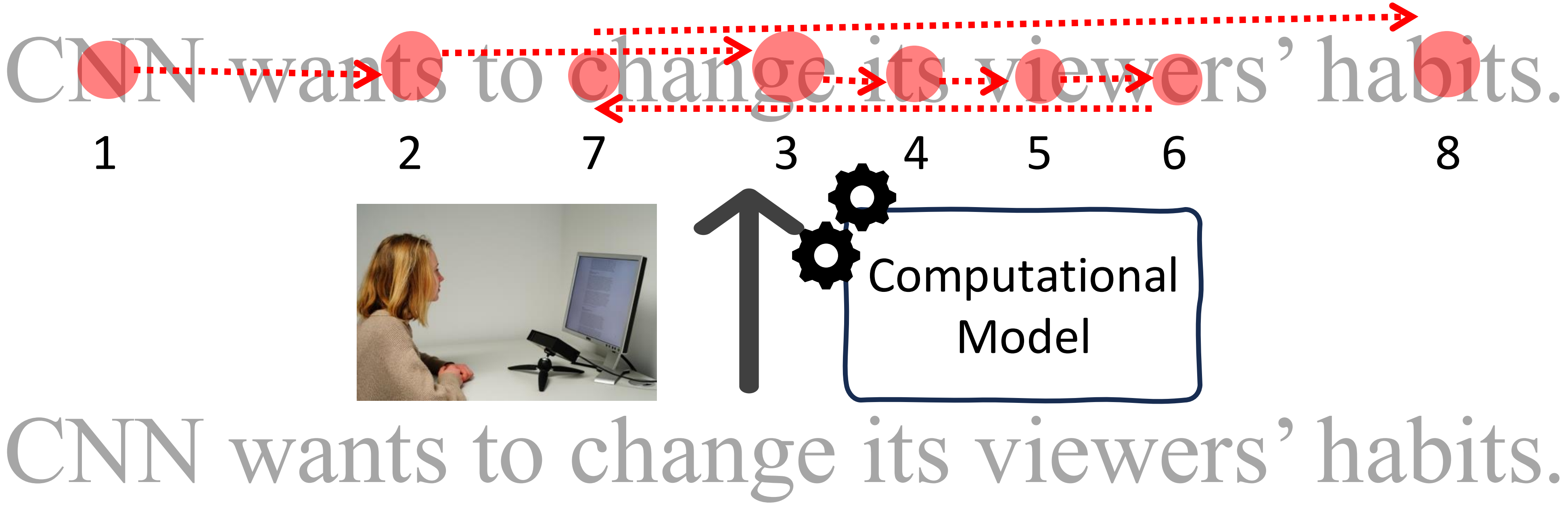
[Malmaud et al. (2020)]

How to combine eye movements and text?



Additional challenges

- Human data is scarce
  - ➢ Scale with synthetic eye movements

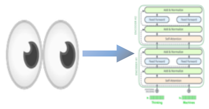


Additional challenges

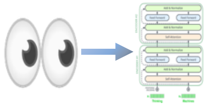
- Human data is scarce
  - ➢ Scale with synthetic eye movements
- In reality, subwordtokens instead of words
  - ➢ Need to decide how to match representations

CNN wants to change its viewers' habits.

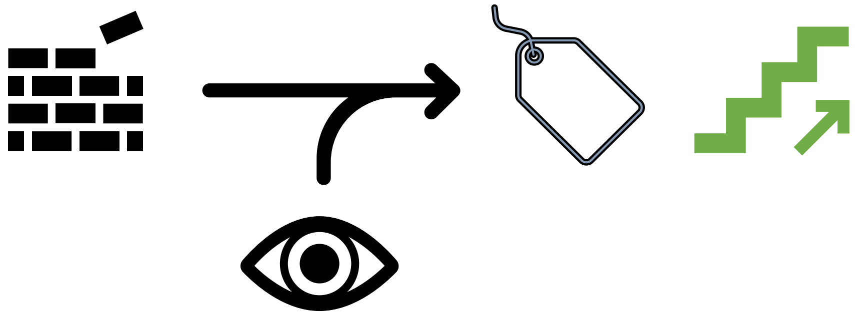
Discussion



- Performance improvements are typically modest
- Hard to beat scale!
  - Currently no convincing example of large and robust improvements for a state-of-the-art LLM
- Possible directions forward
  - Better modeling
  - More human data
  - Higher quality synthetic eye movements
  - Low resource scenarios
  - Multilingual approaches

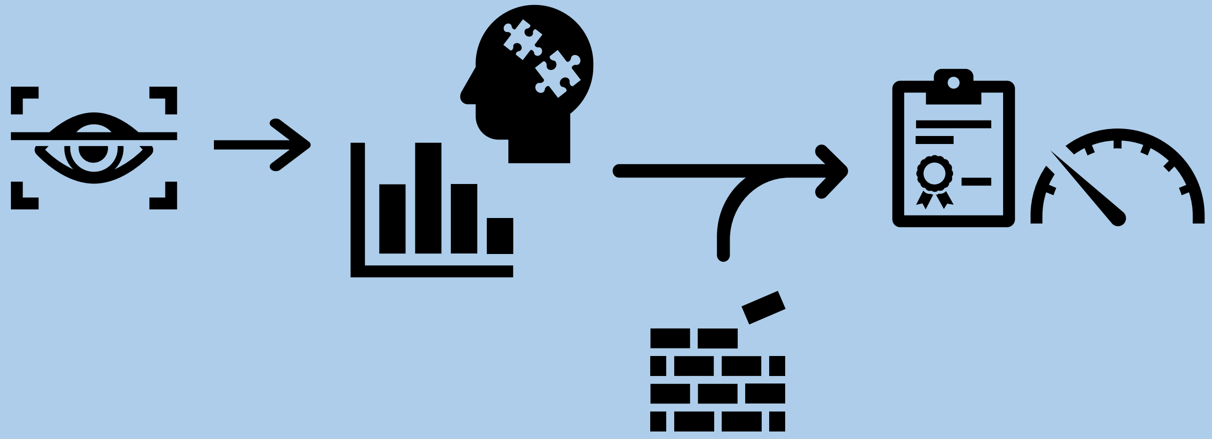
Uses of Eye Tracking in NLP 

Modeling



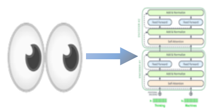
Eye movements can enhance the performance of NLP models

Evaluation



Eye movements for evaluating the performance of NLP models

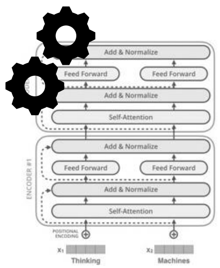
Evaluating Task Performance

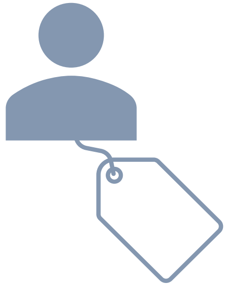


Offline Human Reference

Captures a behavioral **end product** of linguistic processing

Output

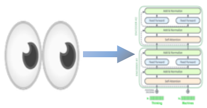




Annotation
Summary
Translation
Preference
...

CNN wants to change its viewers' habits.

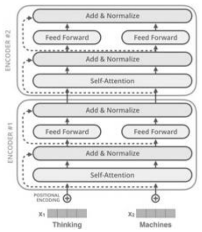
Evaluating Task Performance



Online Human Behavior

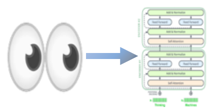
Captures language processing in **real time**

Output

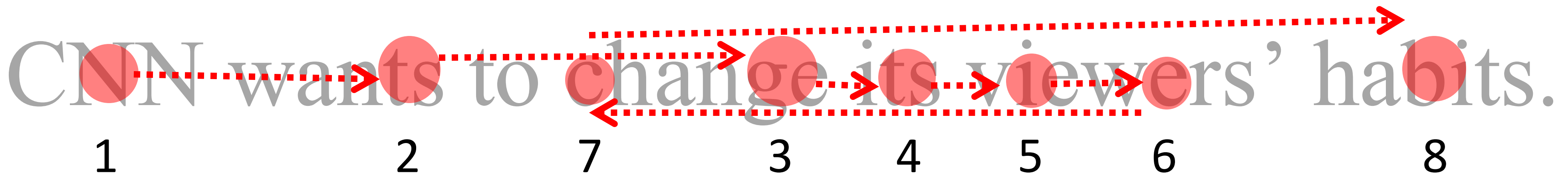


CNN wants to change its viewers' habits.

How do people read?



Eye Mind Assumption: "… *there is no appreciable lag between what is fixated and what is processed.*" Just & Carpenter, 1980



Tight correspondence between eye movements and linguistic processing

Evaluating Task Performance

Online Human Behavior

Captures language processing in real time

Machine translation [Doherty et al. (2010)](#) [Sajjad et al. (2016)](#)

Summarization [Ikhwantri et al. (2024)](#)

Readability [Gruteke Klein et al. (2025)](#)

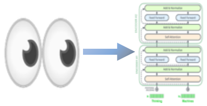
Example Task - ARA

Automatic Readability Assessment (ARA):

Scoring the difficulty level of a text

- Popular task in NLP
- Over 100 years of research
- Hundreds of papers, dozens of measures and systems
- Many real-world applications

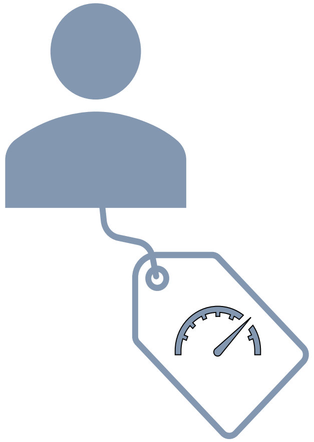
Automatic Readability Assessment



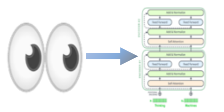
Standard evaluation methods:

Reading Comprehension Performance

Human Labeling



Automatic Readability Assessment



Reader's Comprehension

- Traditional readability measures fitted to reading comprehension data
- Harder to answer reading comprehension questions → the less readable the text.

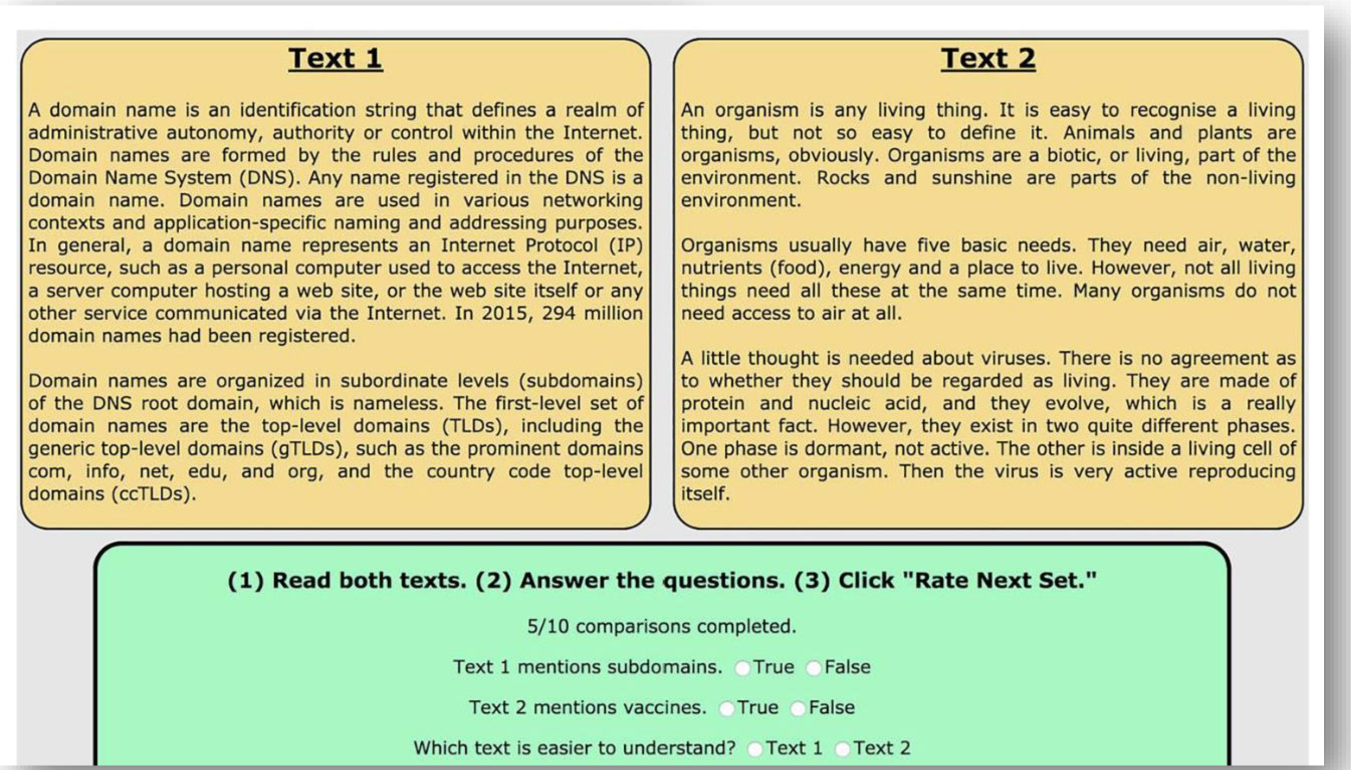
Automatic Readability Assessment

Reader's Comprehension ≠ Text's Readability

- Possible, but not necessary a byproduct of readability
- Depends on the difficulty of the questions
- Hard to estimate reliably

Automatic Readability Assessment

Human labeling → very challenging task



Text 1

A domain name is an identification string that defines a realm of administrative autonomy, authority or control within the Internet. Domain names are formed by the rules and procedures of the Domain Name System (DNS). Any name registered in the DNS is a domain name. Domain names are used in various networking contexts and application-specific naming and addressing purposes. In general, a domain name represents an Internet Protocol (IP) resource, such as a personal computer used to access the Internet, a server computer hosting a web site, or the web site itself or any other service communicated via the Internet. In 2015, 294 million domain names had been registered.

Domain names are organized in subordinate levels (subdomains) of the DNS root domain, which is nameless. The first-level set of domain names are the top-level domains (TLDs), including the generic top-level domains (gTLDs), such as the prominent domains com, info, net, edu, and org, and the country code top-level domains (ccTLDs).

Text 2

An organism is any living thing. It is easy to recognise a living thing, but not so easy to define it. Animals and plants are organisms, obviously. Organisms are a biotic, or living, part of the environment. Rocks and sunshine are parts of the non-living environment.

Organisms usually have five basic needs. They need air, water, nutrients (food), energy and a place to live. However, not all living things need all these at the same time. Many organisms do not need access to air at all.

A little thought is needed about viruses. There is no agreement as to whether they should be regarded as living. They are made of protein and nucleic acid, and they evolve, which is a really important fact. However, they exist in two quite different phases. One phase is dormant, not active. The other is inside a living cell of some other organism. Then the virus is very active reproducing itself.

(1) Read both texts. (2) Answer the questions. (3) Click "Rate Next Set."

5/10 comparisons completed.

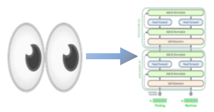
Text 1 mentions subdomains. ○ True ○ False

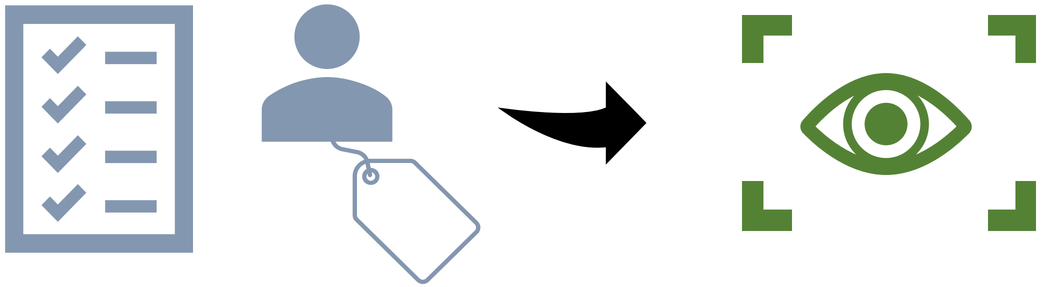
Text 2 mentions vaccines. ○ True ○ False

Which text is easier to understand? ○ Text 1 ○ Text 2

[Crossley et al. (2019)]

Automatic Readability Assessment

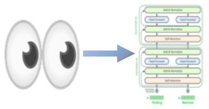


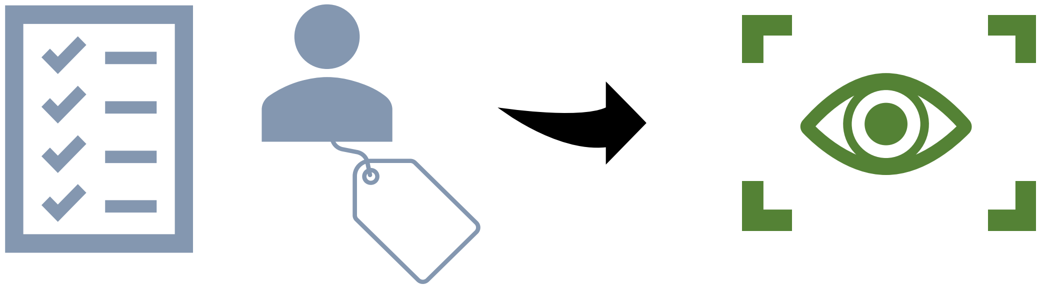


From reading **comprehension** performance and human **labeling** of text difficulty to **cognitive evaluation** of **reading ease** using eye tracking

[Gruteke Klein et al. (2025)](#)

Automatic Readability Assessment





Reading Speed

Average Total Fixation Duration

Skip Rate

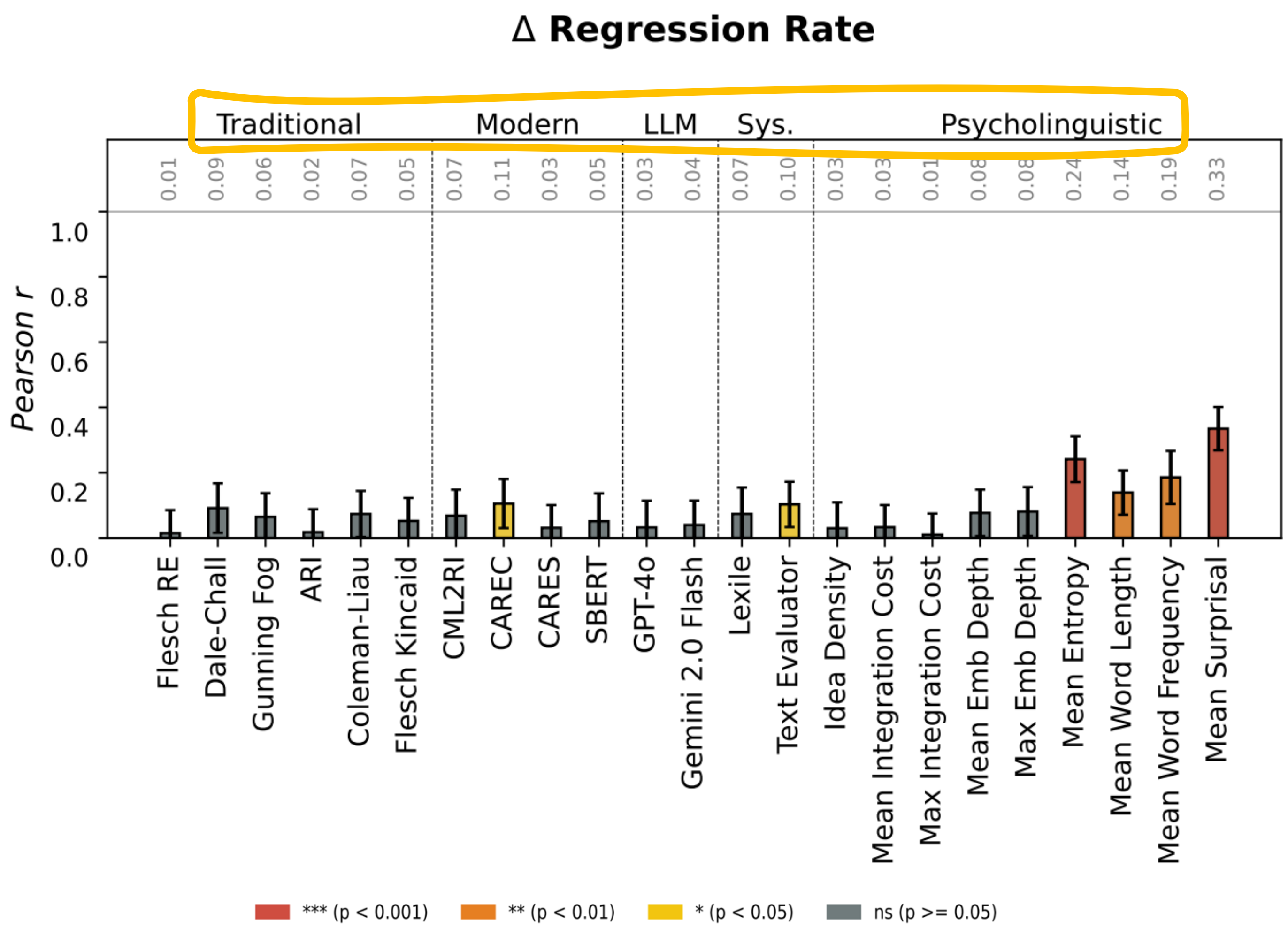
Regression Rate

Evaluation of readability measure M is the Pearson r of:

$$RT \sim Score_M$$

Gruteke Klein et al. (2025)

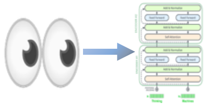
Automatic Readability Assessment



Low correlation between existing measures and reading ease

Gruteke Klein et al. (2025)

Evaluating Task Performance



Online Human Behavior

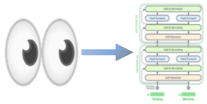
Captures language processing in real time

Machine translation Doherty et al. (2010)
Sajjad et al. (2016)

Summarization Ikhwantri et al. (2024)

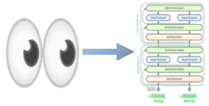
Readability Gruteke Klein et al. (2025)

Discussion: Eye Movements for Evaluating NLP Models

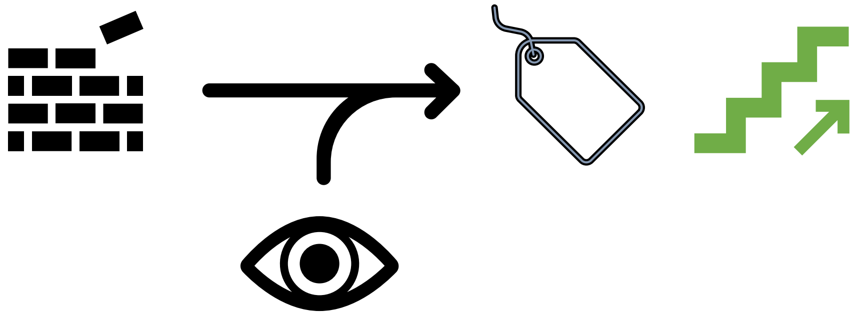


Current uses are limited

- ➢Better use of existing eye tracking data
- ➢Collect more eye tracking data for specific NLP tasks
- ➢Large scale evaluations with synthetic eye movements

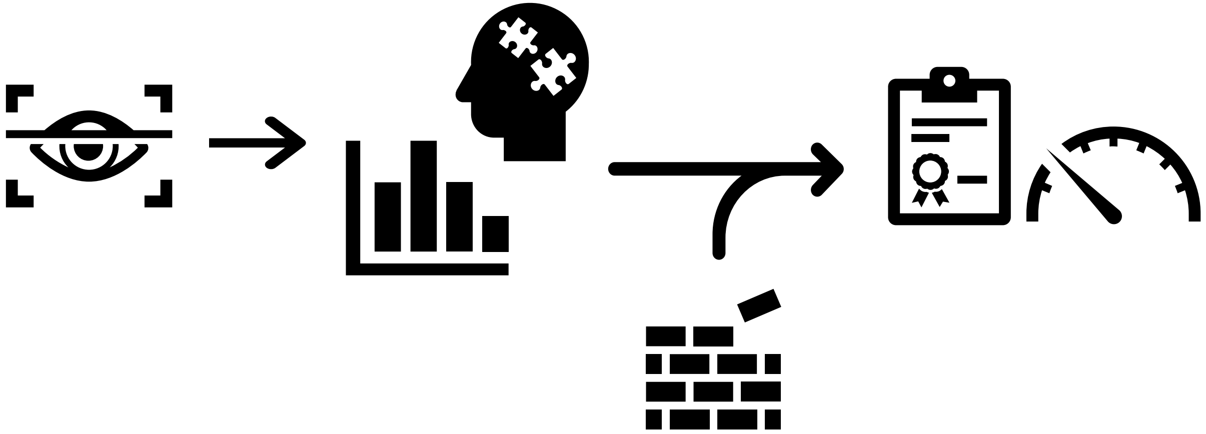
Uses of Eye Tracking in NLP 

Modeling



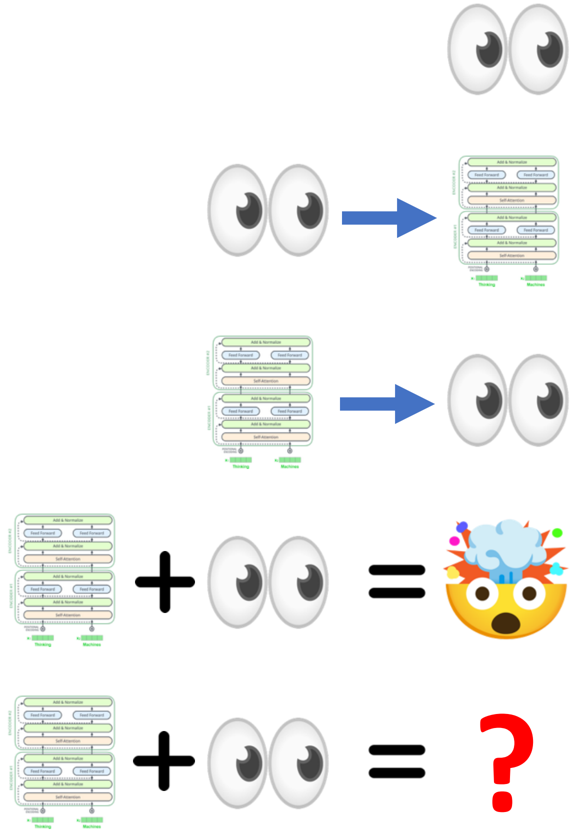
Eye movements can enhance the performance of NLP models

Evaluation



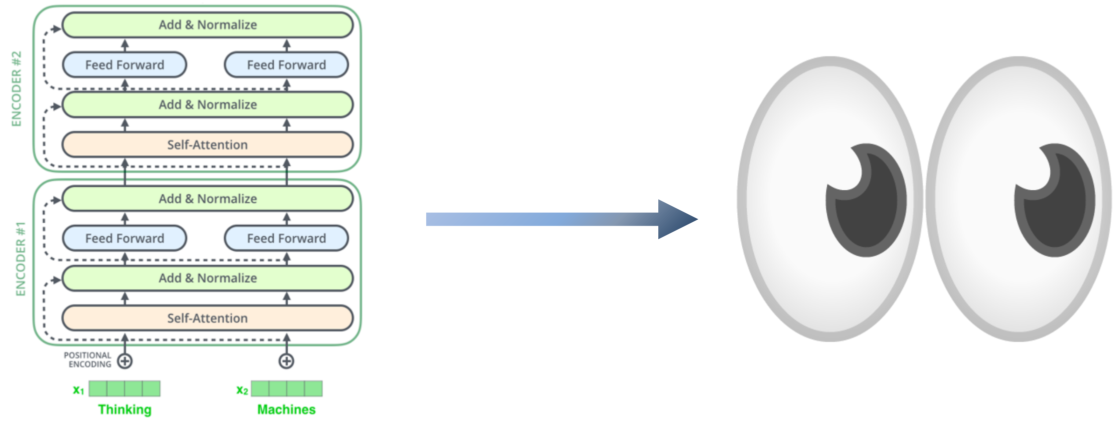
Eye movements for evaluating the performance of NLP models

Tutorial Outline

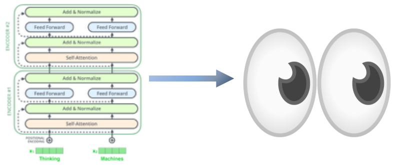


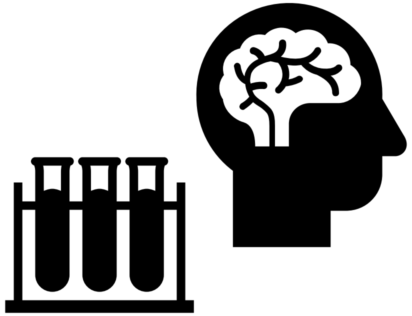
1. Introduction to eye tracking
2. Uses of eye tracking in NLP
3. **NLP for eye movement and cognitive modeling**
4. New human centered applications
5. Outlook and future directions

Uses of NLP in Modeling Human Language Processing and Eye Movements



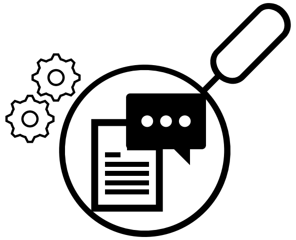
Uses of NLP in Modeling Eye Movements and Human Language Processing



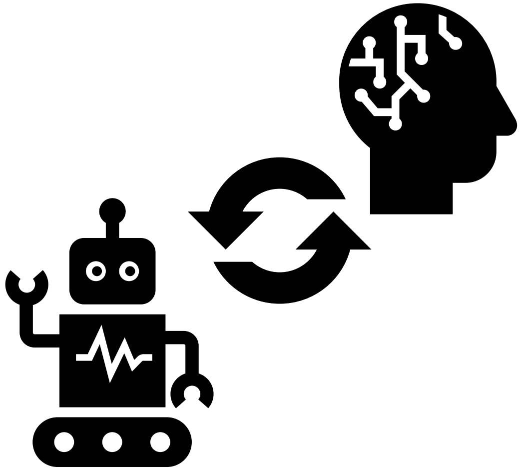


Testing Psycholinguistic Theories

Representations

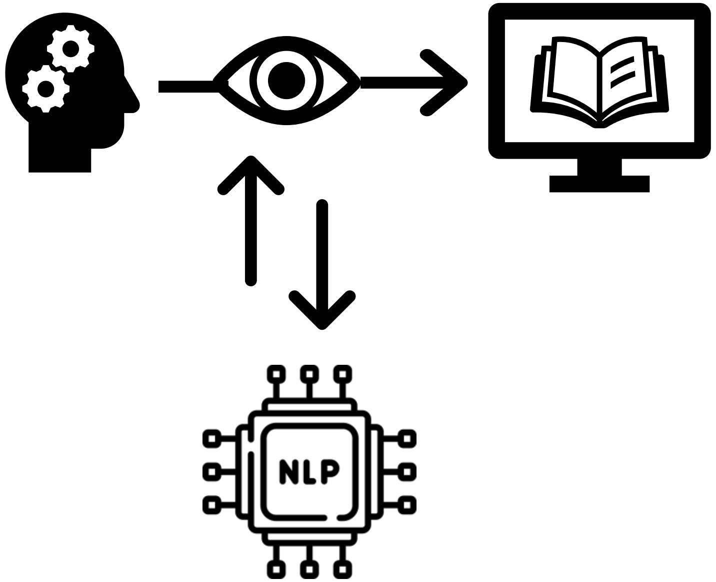


Linguistic quantities

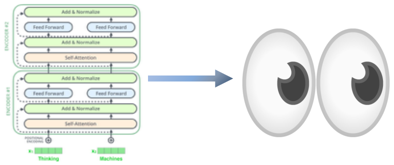


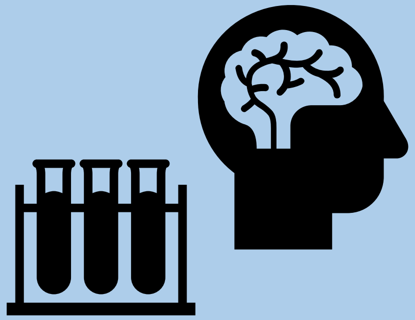
Testing LLM alignment with human language processing

NLP for modeling eye movements in reading

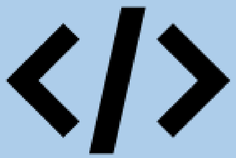


Uses of NLP in Modeling Eye Movements and Human Language Processing

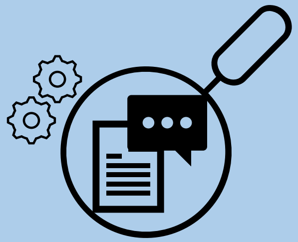




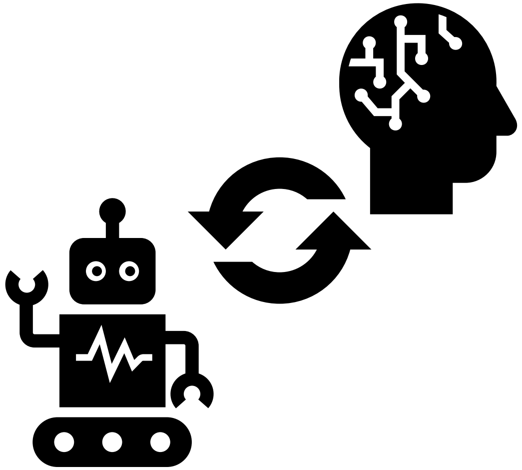
Testing Psycholinguistic Theories



Representations

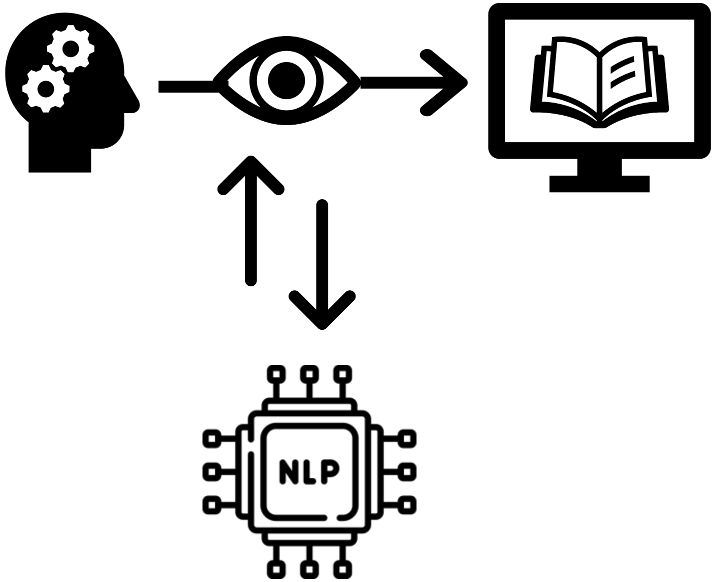


Linguistic quantities

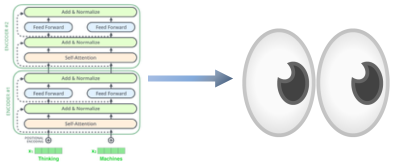


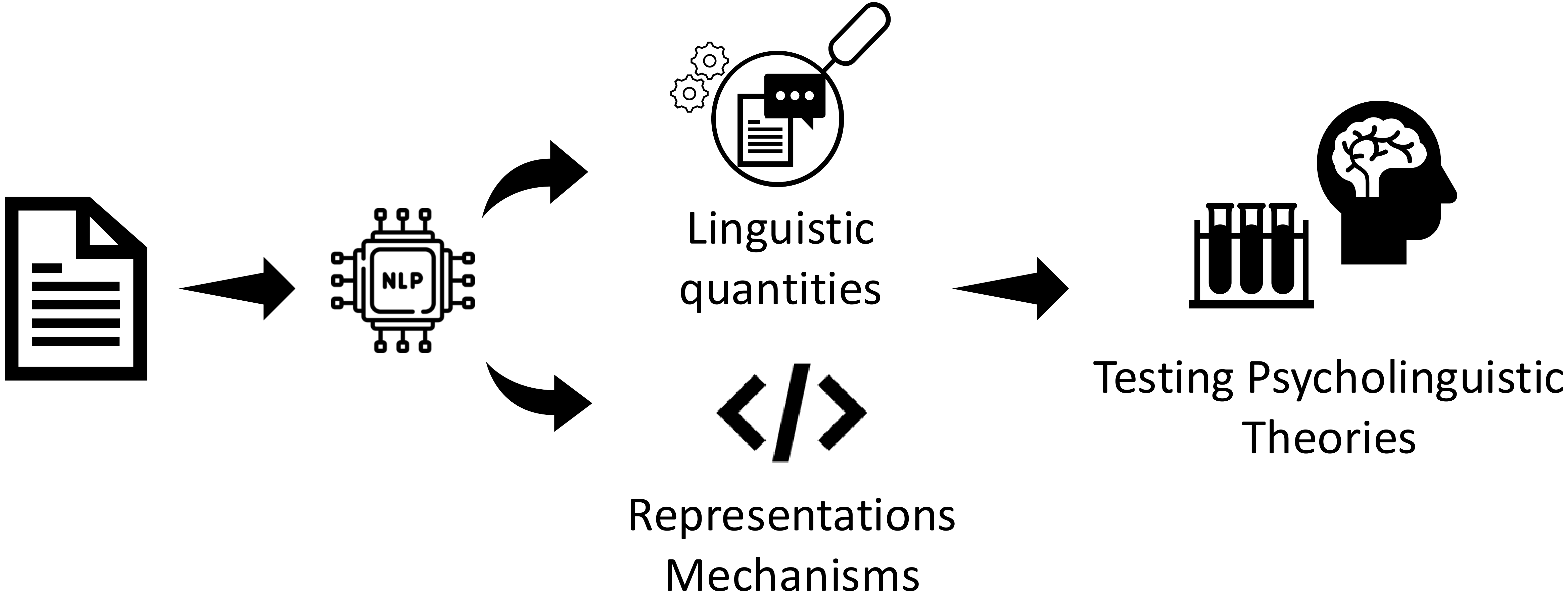
Testing LLM alignment with human language processing

NLP for modeling eye movements in reading

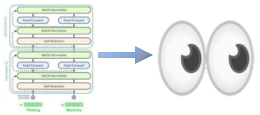


NLP for Extracting Linguistic Quantities and Representations





NLP for Extracting Linguistic Quantities and Representations



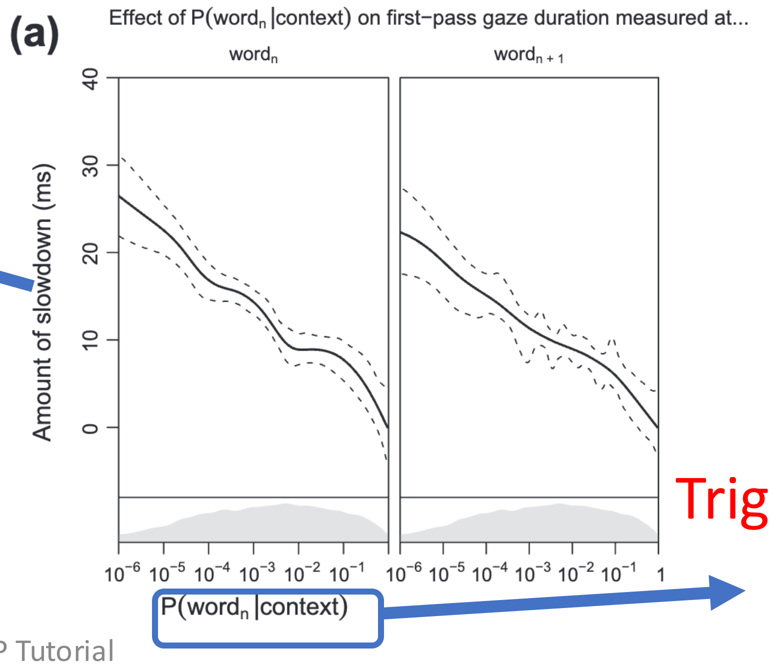
- NLP for broad coverage stimulus annotations
  - Surprisal, entropy
  - Syntactic structure
  - Semantic relations
  - Word embeddings
- Annotations (or derived quantities) are central for testing psycholinguistic theories of language processing
  - Surprisal theory Hale (2001), Levy (2008)
  - Dependency locality theory (DLT) Gibson (1998), Gibson (2000)
  - Uniform information density (UID) Levy & Jaeger (2007)
  - Cue-based retrieval (ACT-R) Lewis & Vasishth (2005), Engelmann et al. (2019)

NLP for Extracting Linguistic Quantities and Representations

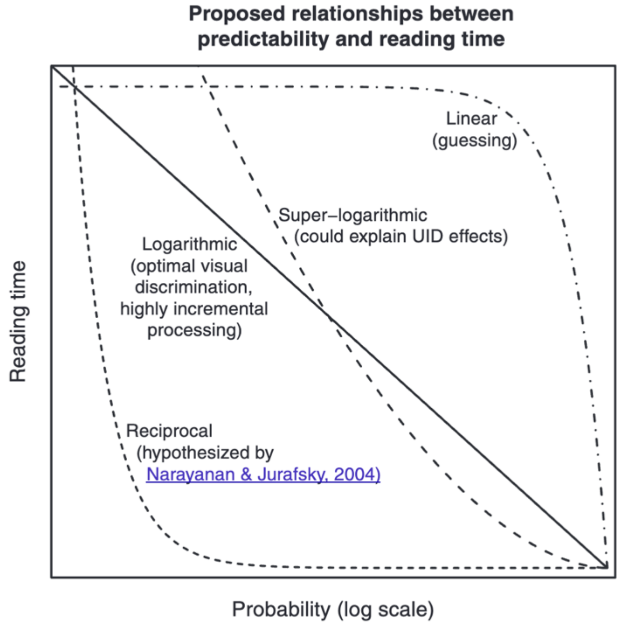
Smith and Levy (2013) testing predictions of Surprisal Theory

- Reading times (Dundee) as proxy for processing difficulty
- **LM based surprisal** as proxy for word predictability

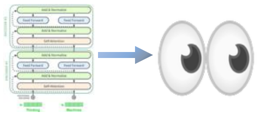
Gaze duration



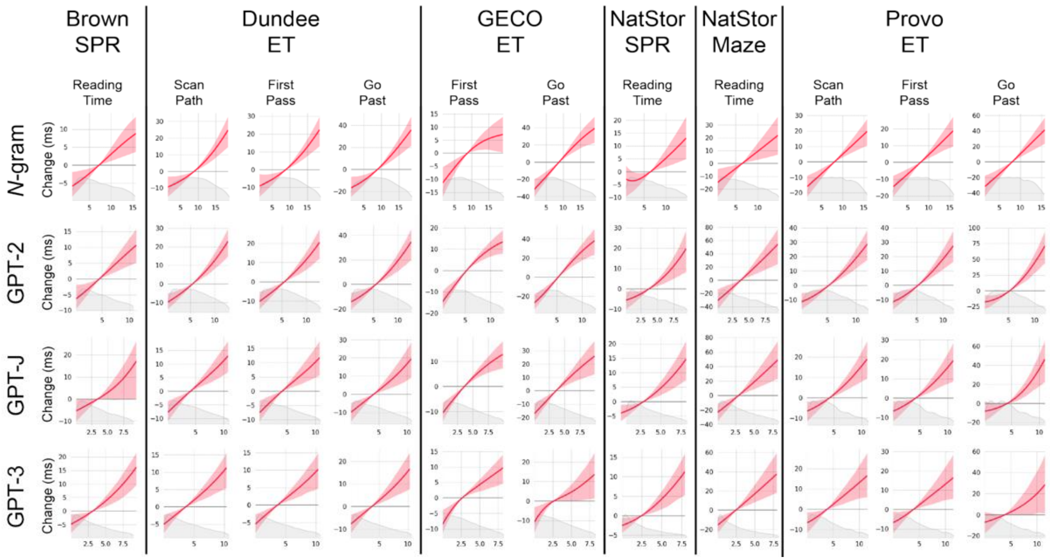
Trigram LM



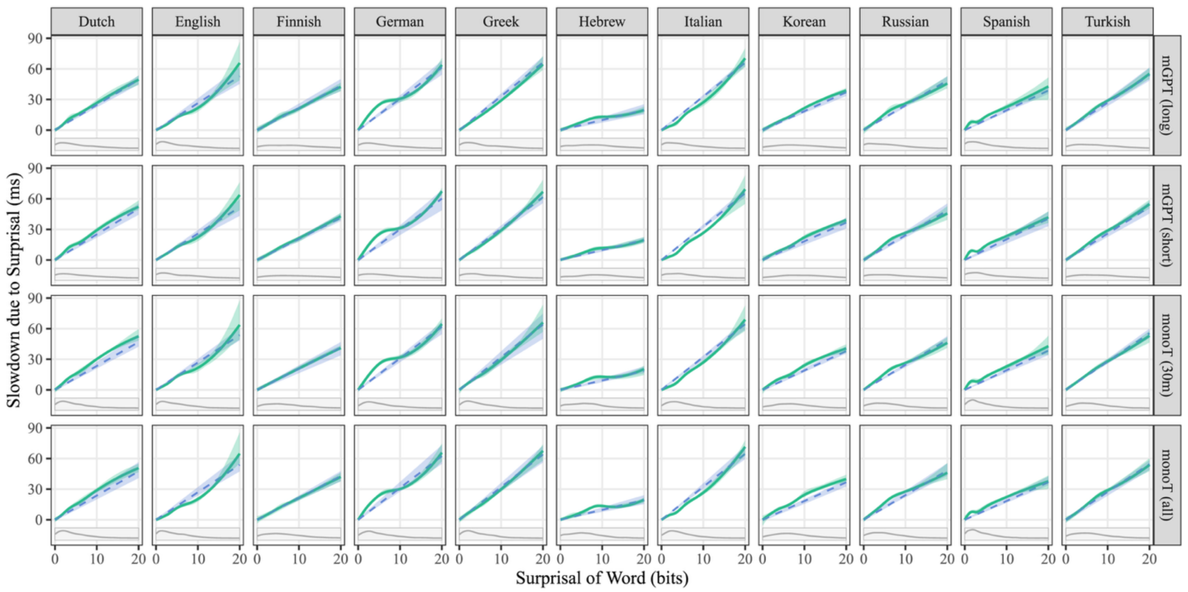
NLP for Extracting Linguistic Quantities and Representations



Linearity holds for different LLMs Shain et al. (2023) and across languages Wilcox et al (2023)



Shain et al. (2023)

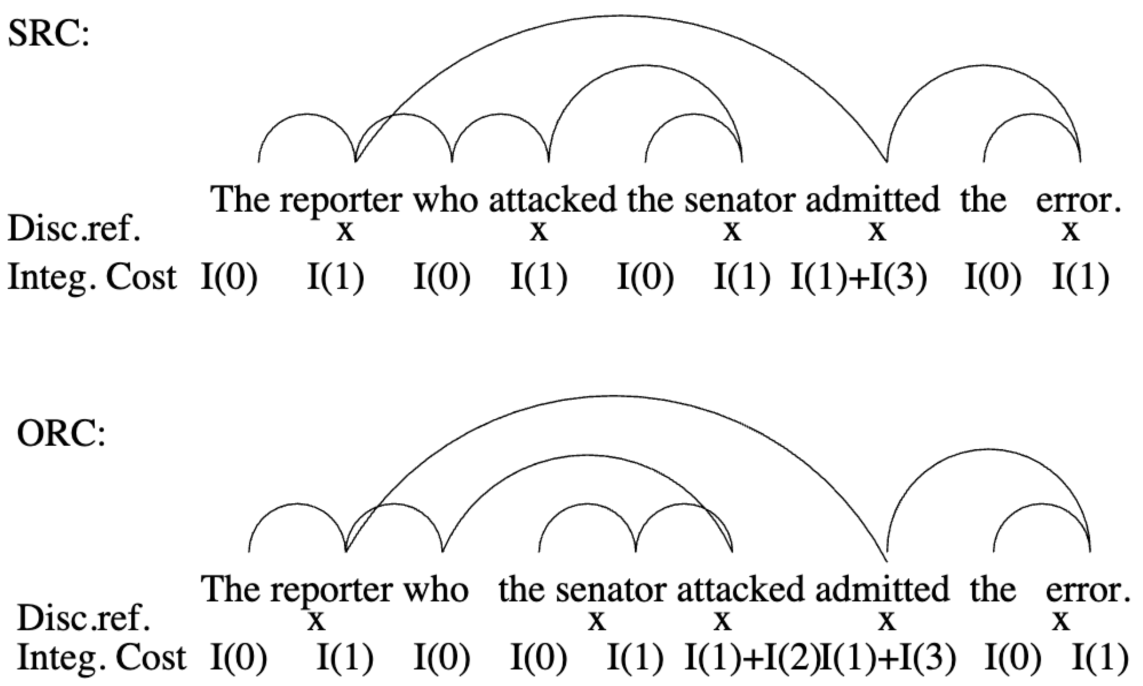


MECO data, Wilcox et al (2023)

NLP for Extracting Linguistic Quantities and Representations

Testing predictions of DLT and Surprisal Theory [Demberg and Keller (2008)]

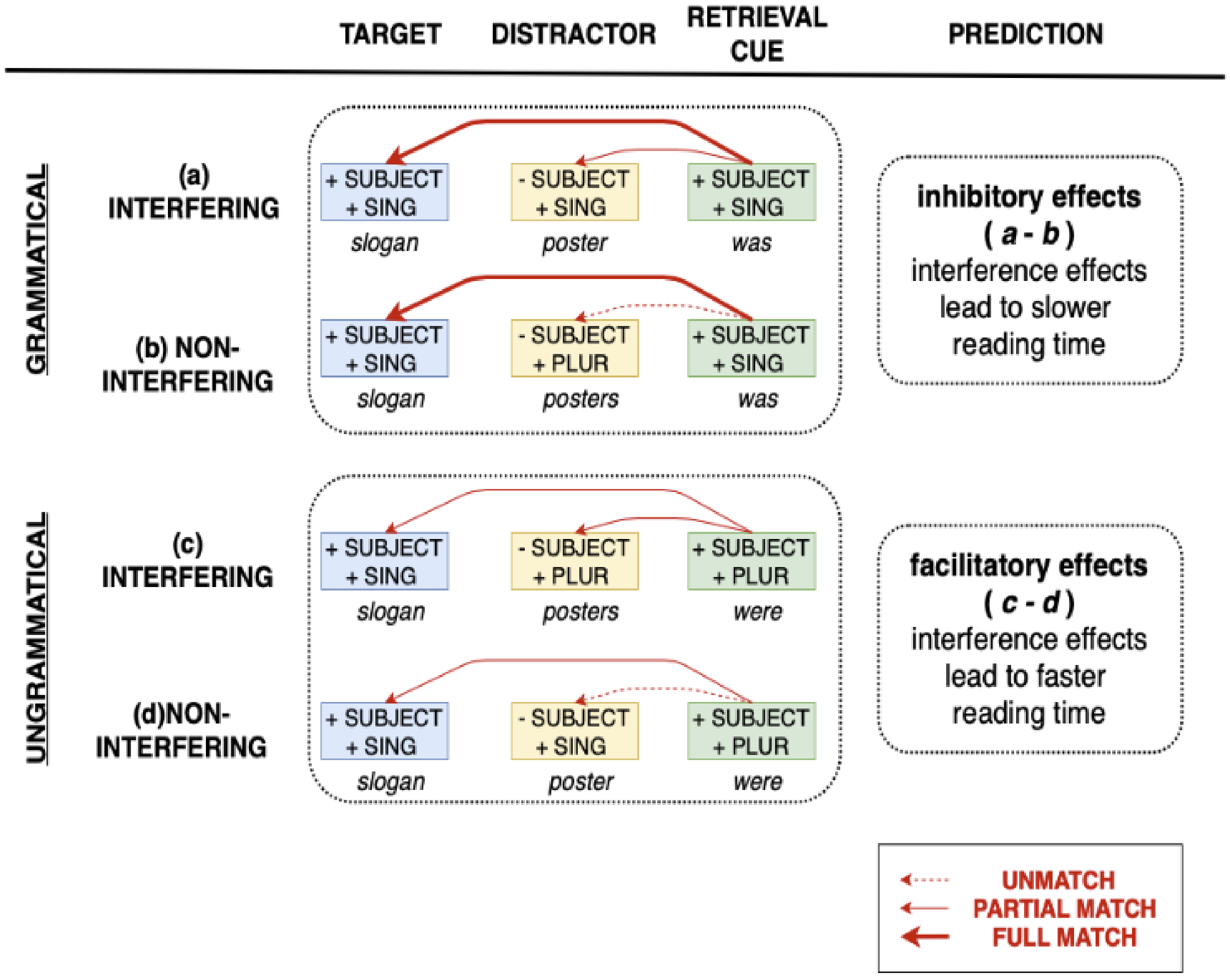
- Reading times (Dundee) as proxy for processing difficulty
- **LM based surprisal** as proxy for word predictability
- **Dependency parsing** for obtaining sentence structure



NLP for Extracting Linguistic Quantities and Representations

Controlled experiments with GPT2 surprisal for testing surprisal and interference based explanations to agreement phenomena

[Ryu and Lewis (2021)](#)



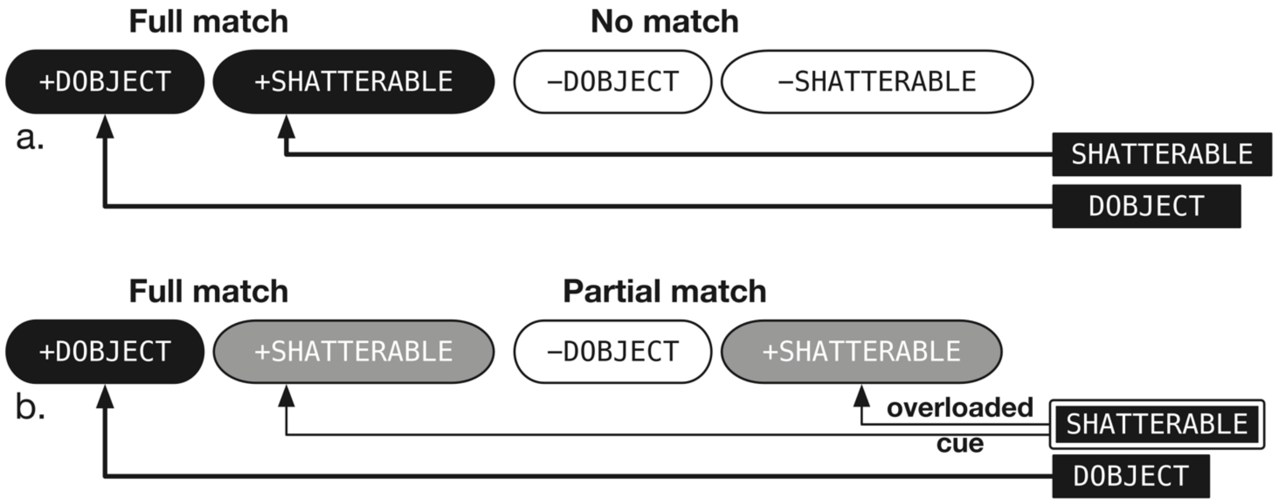
NLP for Extracting Linguistic Quantities and Representations

a) … the **plate** that the butler with the **tie** accidentally **shattered** …

b) … the **plate** that the butler with the **cup** accidentally **shattered** …

Cue-based Retrieval (ACT-R) assumes a content-addressable memory.

Dependency formation: Retrieval cues serve to access relevant chunks (e.g., words) from memory.

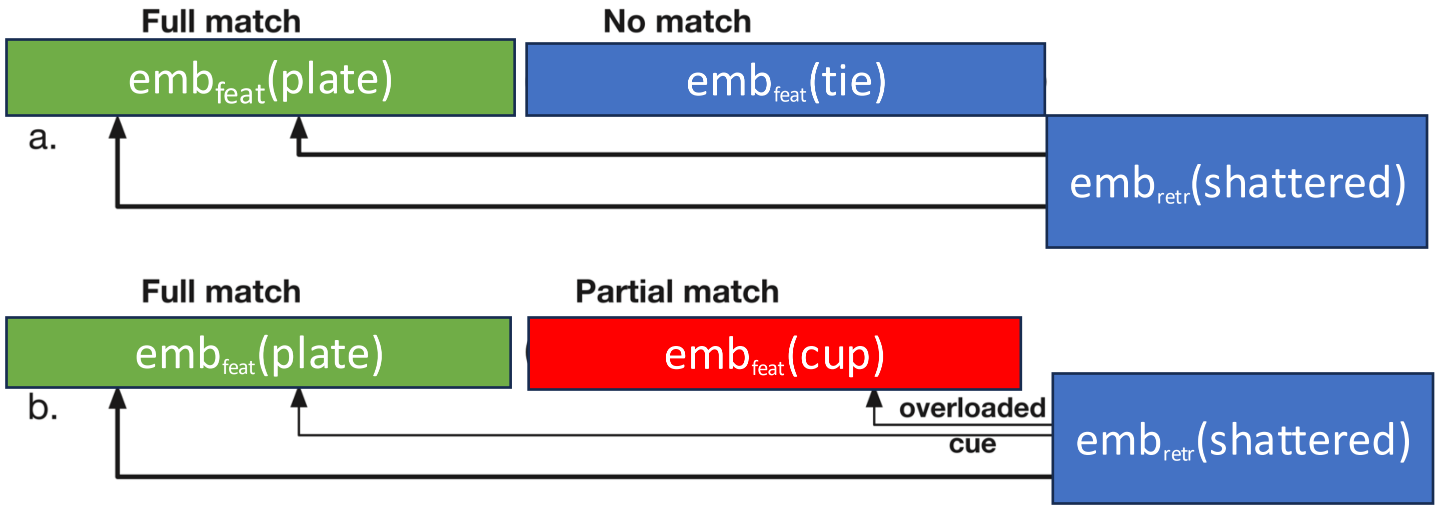


NLP for Extracting Linguistic Quantities and Representations

a) … the **plate** that the butler with the **tie** accidentally **shattered** …

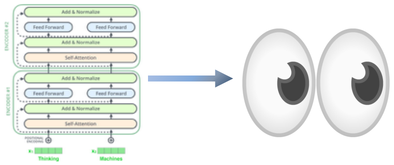
b) … the **plate** that the butler with the **cup** accidentally **shattered** …

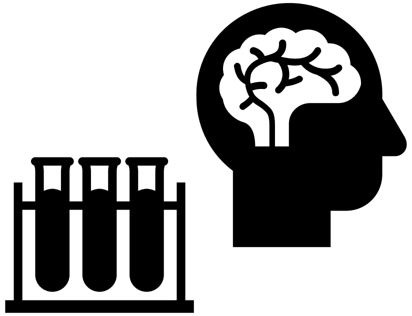
Replacing hand-crafted feature vectors with **word embeddings** as **cognitive representations** for **lexical items** in memory.



Smith and Vasishth (2020)

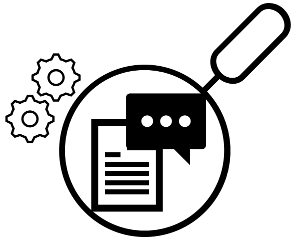
Uses of NLP in Modeling Eye Movements and Human Language Processing



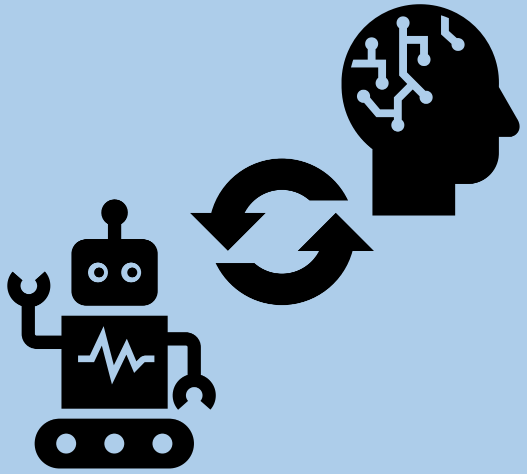


Testing Psycholinguistic Theories

Representations

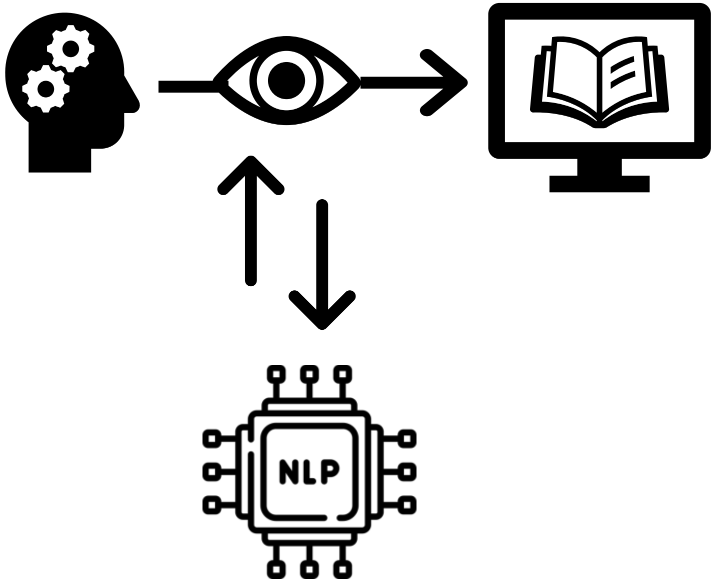


Linguistic quantities

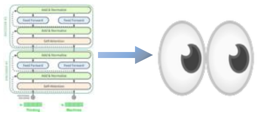


Testing LLM alignment with human language processing

NLP for modeling eye movements in reading



LLM Alignment with Human Reading



- Are LLMs good models of human linguistic processing? If not, how can we make them relevant?
- Big and open area where eye tracking data could play a larger role
- Current directions:
  - Testing alignment of LLMs with reading data
  - Improving alignment with more cognitively plausible architectures

LLM Alignment with Human Reading

Predictive power of LLM-extracted **surprisal** (or other metrics) for human reading times (RTs)

Step 1:
Fit two regression models predicting reading times, **with** and **without surprisal as predictor**

Step 2:
Compute Log-Likelihood (LL) of each model

Step 3:
Predictive Power (PP) of surprisal is the Δ**LL** of the two models

$$\mathcal{M}_{baseline}: RT \sim baseline\_variables \longrightarrow LL(\mathcal{M}_{baseline})$$

$$\mathcal{M}_{surprisal}: RT \sim baseline\_variables \oplus surprisal \longrightarrow LL(\mathcal{M}_{surprisal})$$

$$\longrightarrow \mathrm{PP(surprisal)} \coloneqq \Delta LL(\mathcal{M}_{surprisal}, \mathcal{M}_{baseline}) \coloneqq LL(\mathcal{M}_{surprisal}) - LL(\mathcal{M}_{baseline})$$

LLM Alignment with Human Reading

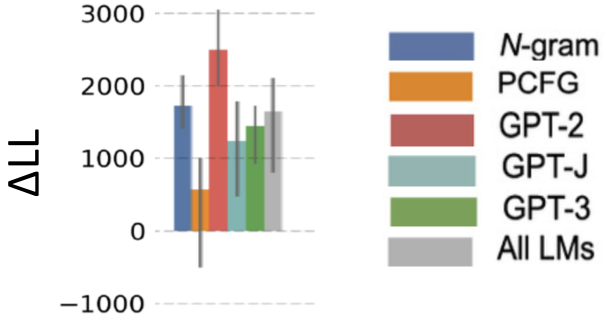
Older Models

Better LM --> surprisal predicts RTs better

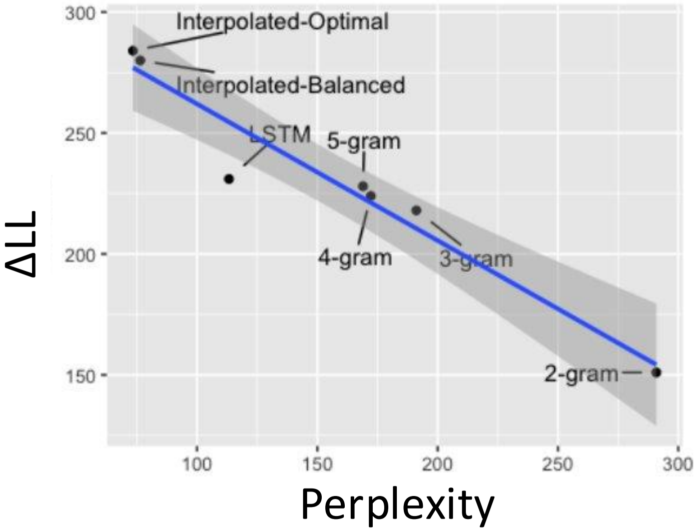
LLMs

Better LM --> surprisal predicts RTs worse

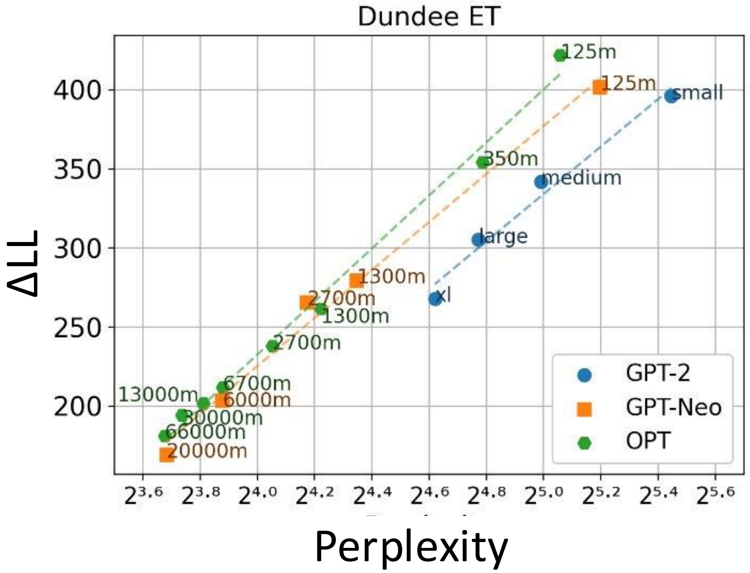
Shain et al. (2023)



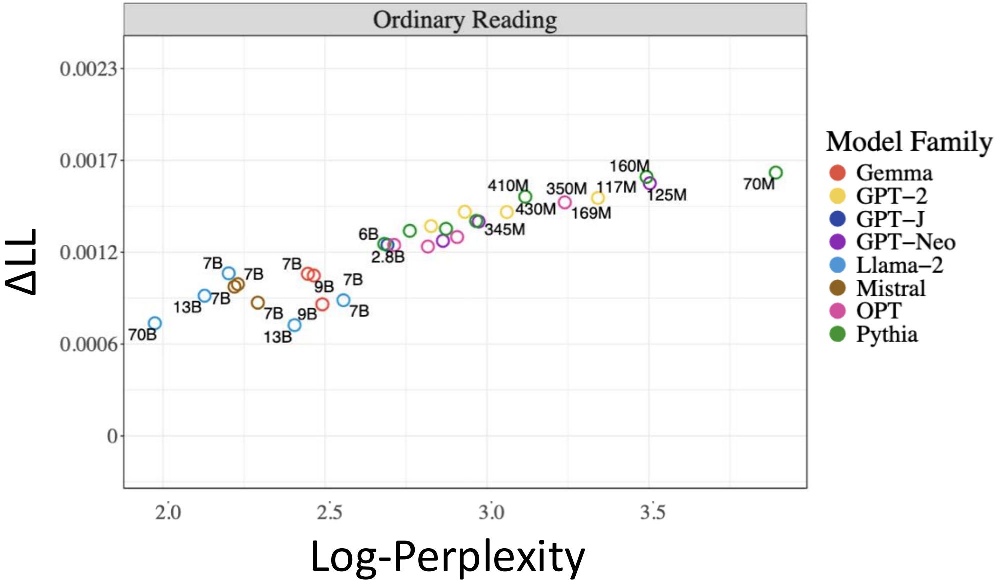
Goodkind and Bicknell (2018)



Oh and Schuler (2023)

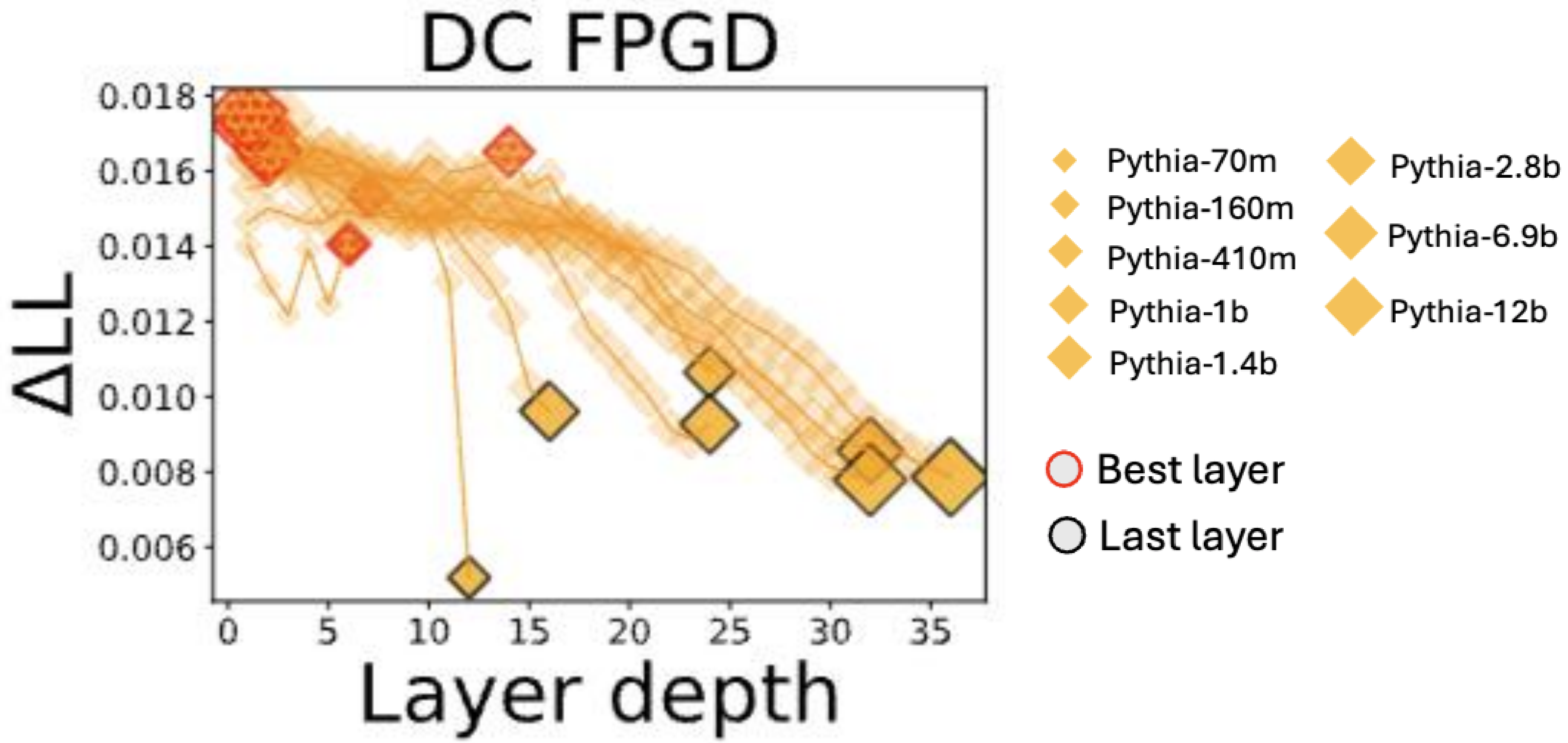


Gruteke Klein et al. (2024)



LLM Alignment with Human Reading

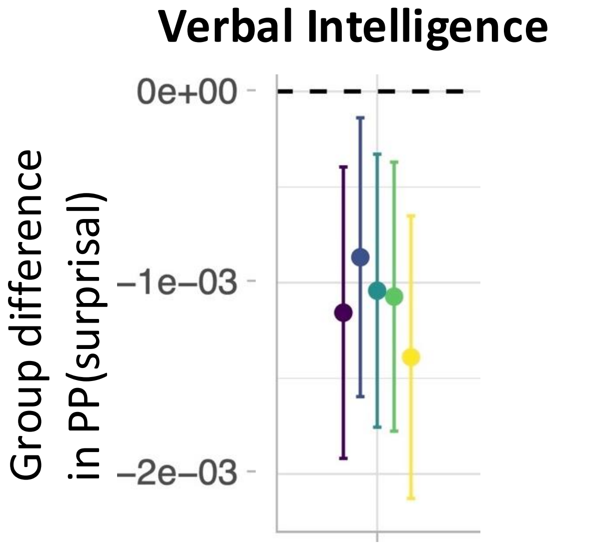
[Kuribayashi et al. (2025)](#) opposite conclusion for intermediate layers

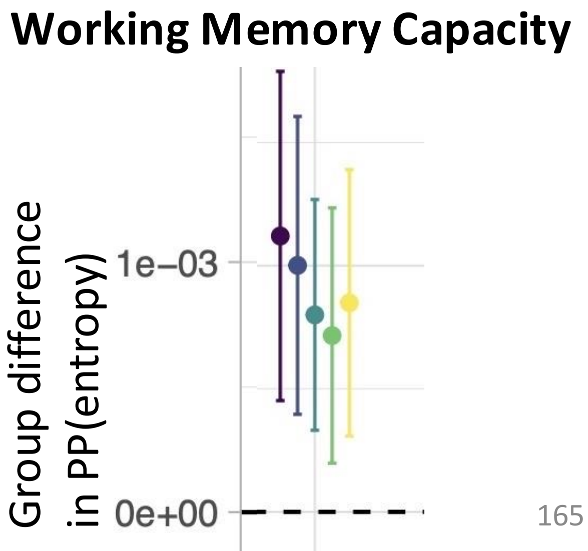


LLM Alignment with Human Reading

Haller et al. (2024) compare the PP on first-pass reading time of LM surprisal and LM entropy for **different cognitive groups of readers**.

- PP of surprisal is higher for readers with **lower verbal intelligence**.
- PP of entropy is higher for readers with **higher working memory capacity**.





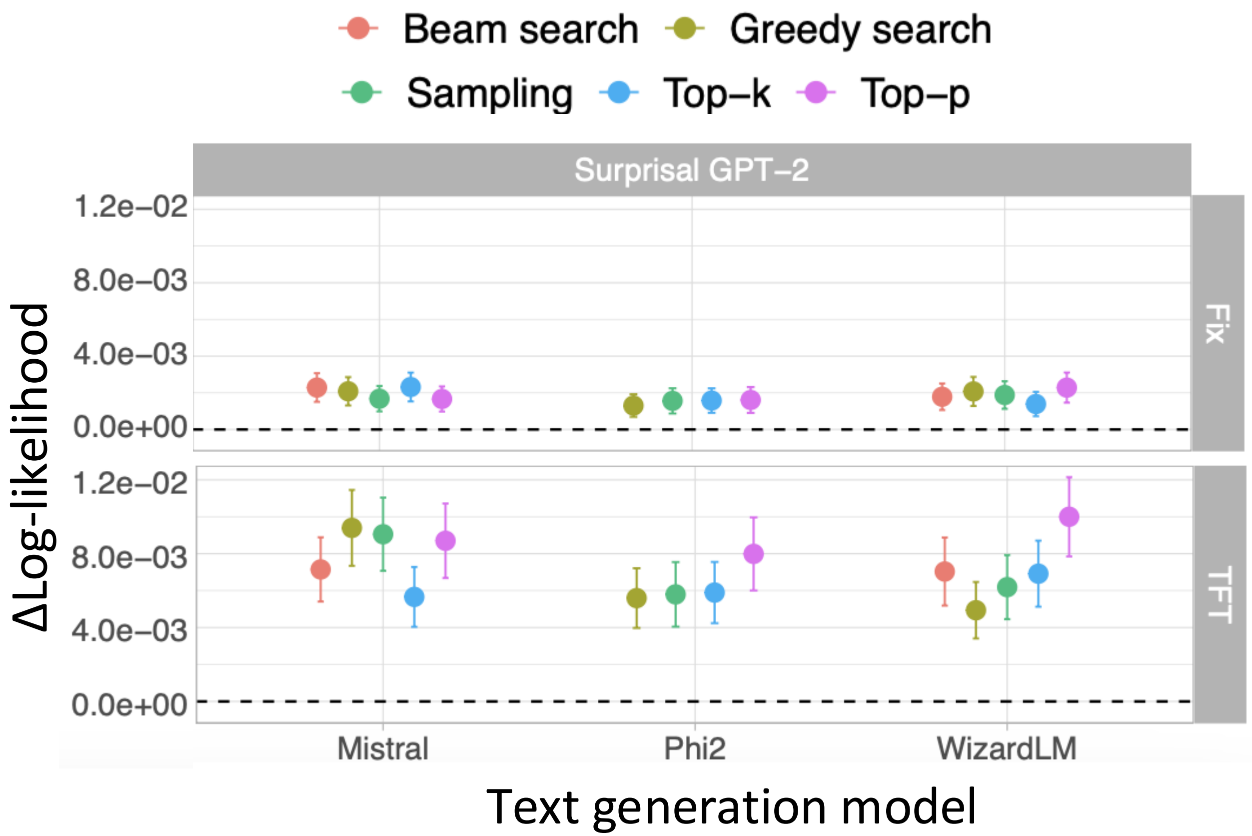
LLM Alignment with Human Reading

Bolliger et al. (2024) Predictive power for reading times varies between texts generated by different

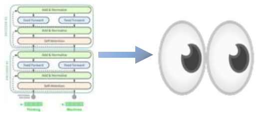
- models
- decoding strategies

And across different

- **reading measures**



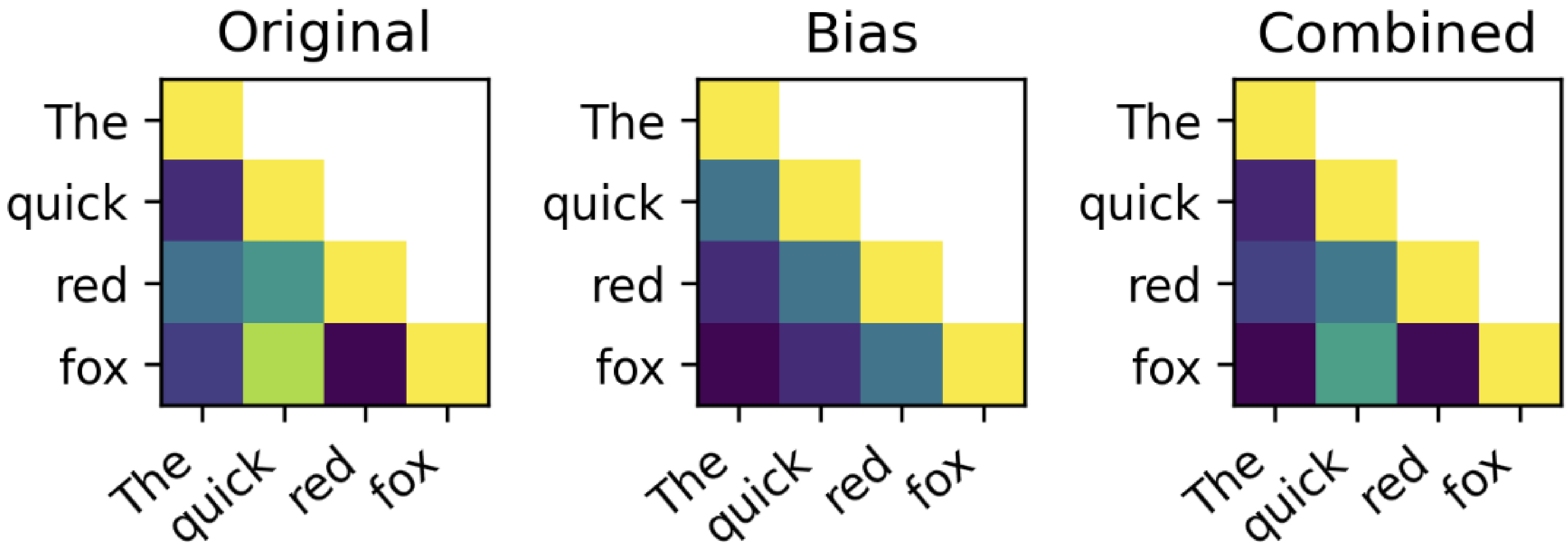
LLM Alignment with Human Reading

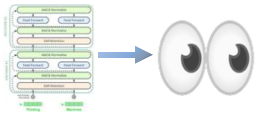


- Rich literature that focuses on
  - Models trained on human scale data
  - Controlled evaluations of targeted linguistic phenomena
- LLMs as a lower bound of what can be learned from the input without postulating innate linguistic knowledge ("poverty of the stimulus")
- Current online evals are primarily on SPR and Maze. E.g. van Schijndel & Linzen (2018), Wilcox et al. (2021)
- Open area for future work with eye tracking data!

Improving Alignment: Adding Cognitive Constraints

- Recency bias for transformer attention de Varda and Marelli (2024), Clark et al (2025)

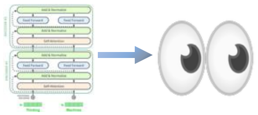




Discussion: Alignment

- Most studies focus on a single reading measure (e.g., Gaze Duration, Total Fixation Duration)
- No clear advantage of eye tracking over other methods
- Similar evaluations are done with cheaper methods that can be (and are) deployed on the web at scale (SPR, Maze)
- Possible directions forward
  - More fine-grained analyses of reading measures to reveal dynamics over time, scanpath prediction
  - Populations with different linguistic knowledge (e.g. L1 vs L2)

Addendum: Parsers as Cognitive Models

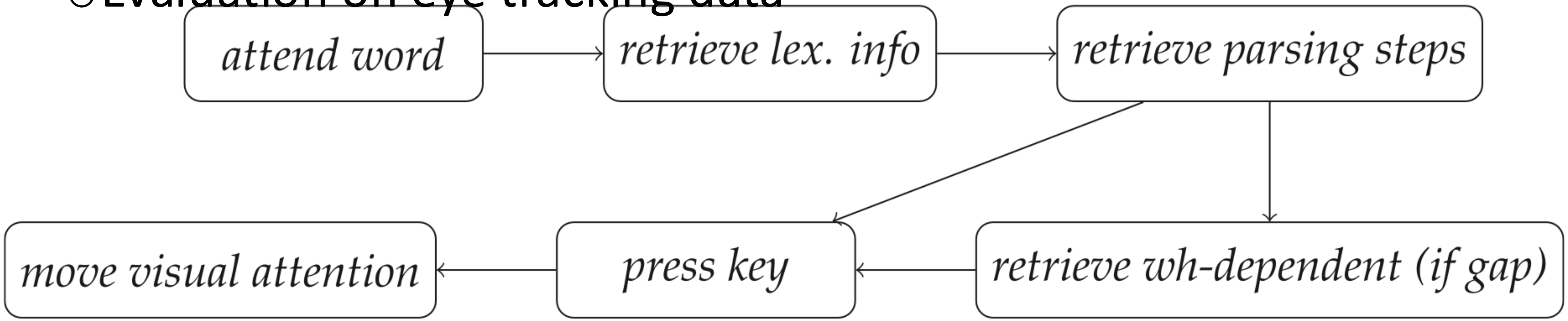


- Long Tradition in computational linguistics (e.g., [Jurafsky, 1996])

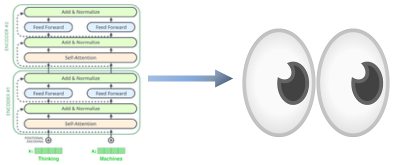
Example: [Dotlačil (2021)]

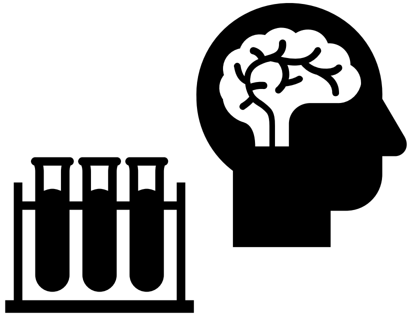
Transition based parser combined with cue-based retrieval (ACT-R)

  - Evaluation on eye tracking data

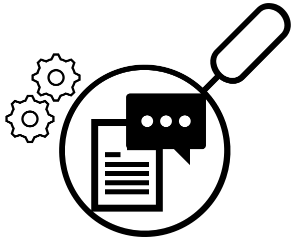


Uses of NLP in Modeling Eye Movements and Human Language Processing



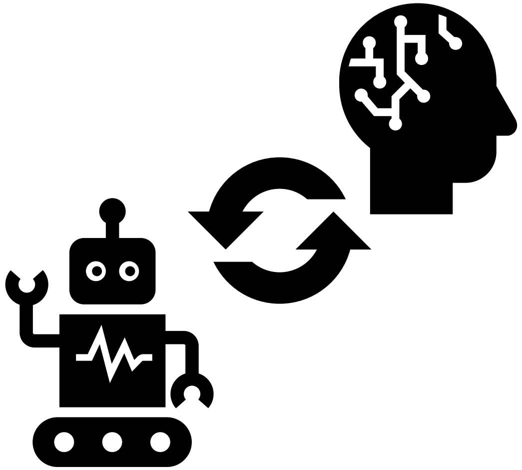


Testing Psycholinguistic Theories



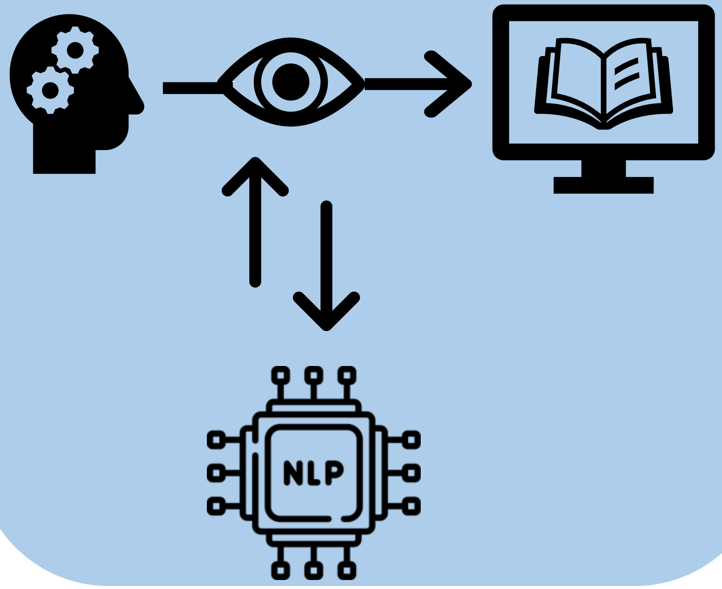
Representations Mechanisms

Linguistic quantities

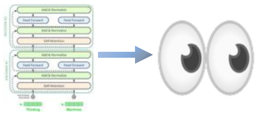


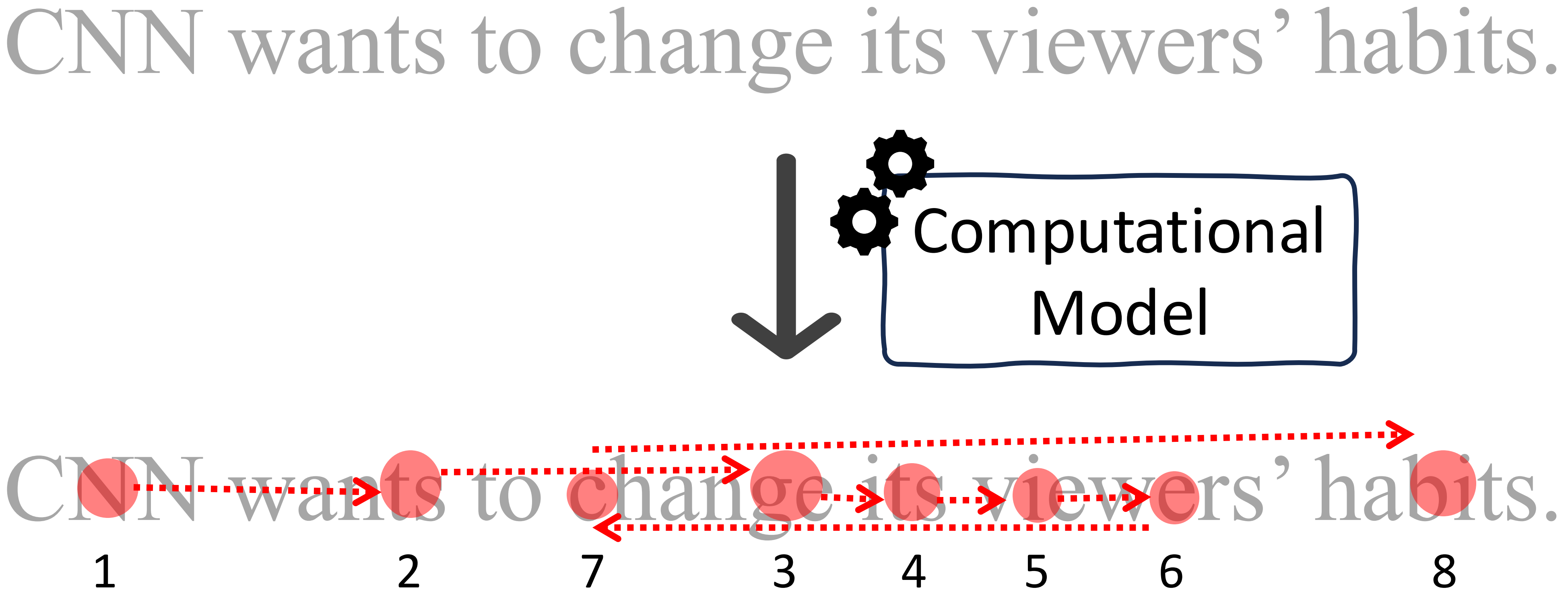
Testing LLM alignment with human language processing

NLP for modeling eye movements in reading

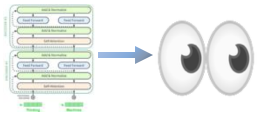


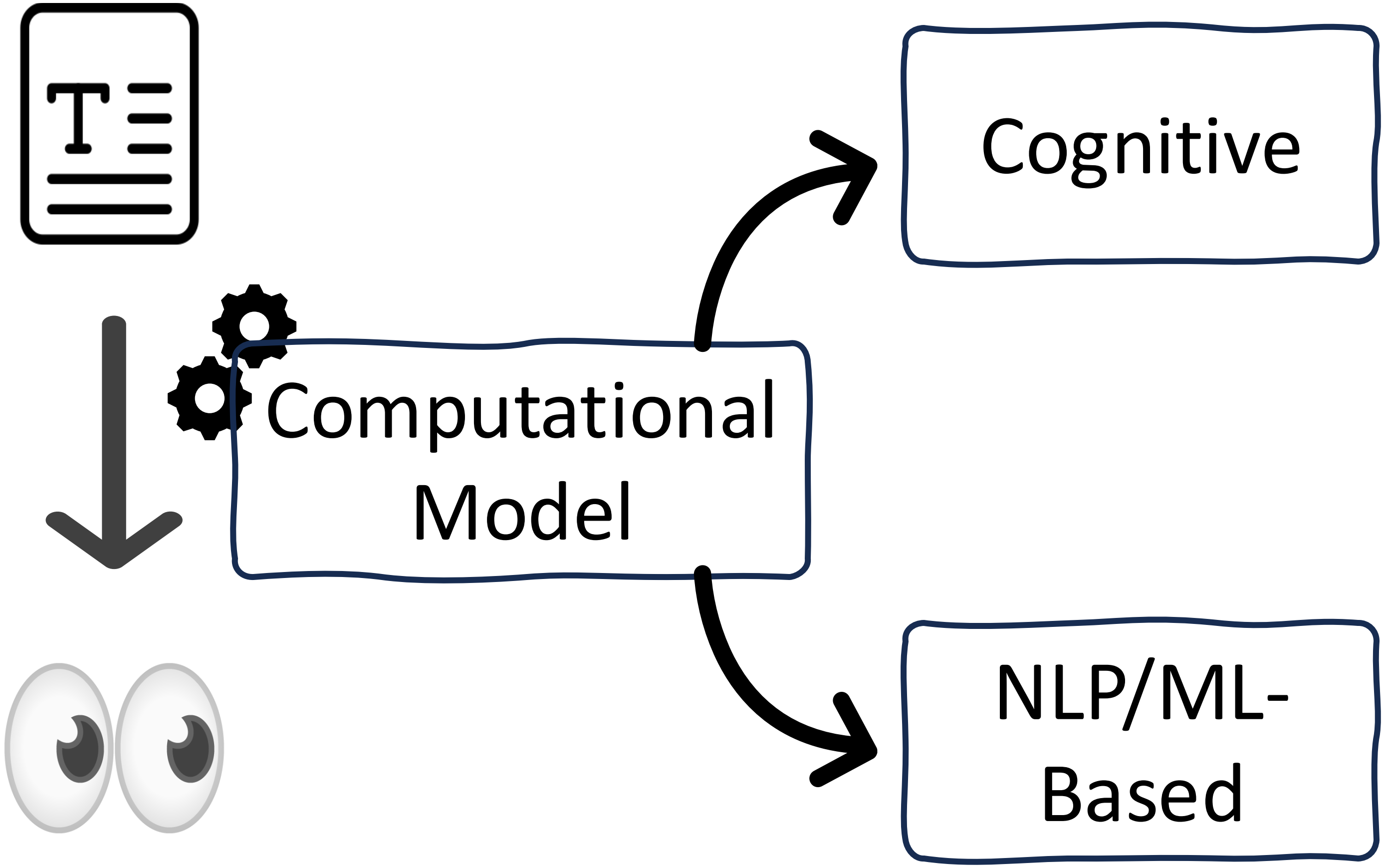
NLP for Modeling Eye Movements in Reading





NLP for Modeling Eye Movements in Reading

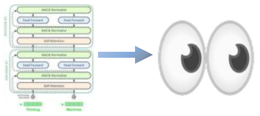




Computationally implemented theories of human reading

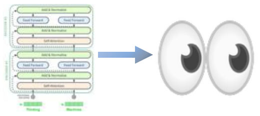
Deep neural networks trained to generate human-like eye movements

NLP for Modeling Eye Movements in Reading



		# Parameters	Interpretable?
Cognitive	Computationally implemented theories of human reading	Few	Most parameters have direct cognitive interpretation
NLP/ML-Based	Deep neural networks trained to generate human-like eye movements	Many	Typically not interpretable

NLP for Modeling Eye Movements in Reading



Examples

Cognitive

Computationally implemented theories of human reading

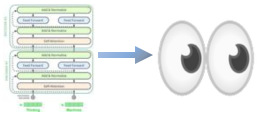
- E-Z Reader Reichle et al. (1998, 2009)
- SWIFT Engbert et al. (2005)
- SEAM Rabe et al. (2024)
- OB1-Reader Snell et al. (2018)

NLP/ML-Based

Deep neural networks trained to generate human-like eye movements

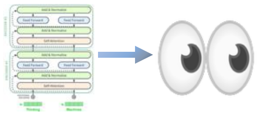
- NEAT Hahn and Keller (2023)
- Eyettention Deng, Reich et al. (2023)
- ScanDL Bolliger et al. (2023, 2025)
- SP-EyeGan Prasse, Reich et al. (2023)

NLP for Modeling Eye Movements in Reading



	Examples	Model
Cognitive	E-Z Reader Reichle et al. (1998, 2009)	Serial attention
	SWIFT Engbert et al. (2005)	Parallel attention
	SEAM Rabe et al. (2024)	Activation-coupled
	OB1-Reader Snell et al. (2018)	Bayesian inference
NLP/ML-Based	NEAT Hahn and Keller (2023)	RNN
	Eyettention Deng, Reich et al. (2023)	Cross-attention
	ScanDL Bolliger et al. (2023, 2025)	Diffusion
	SP-EyeGan Prasse, Reich et al. (2023)	GAN

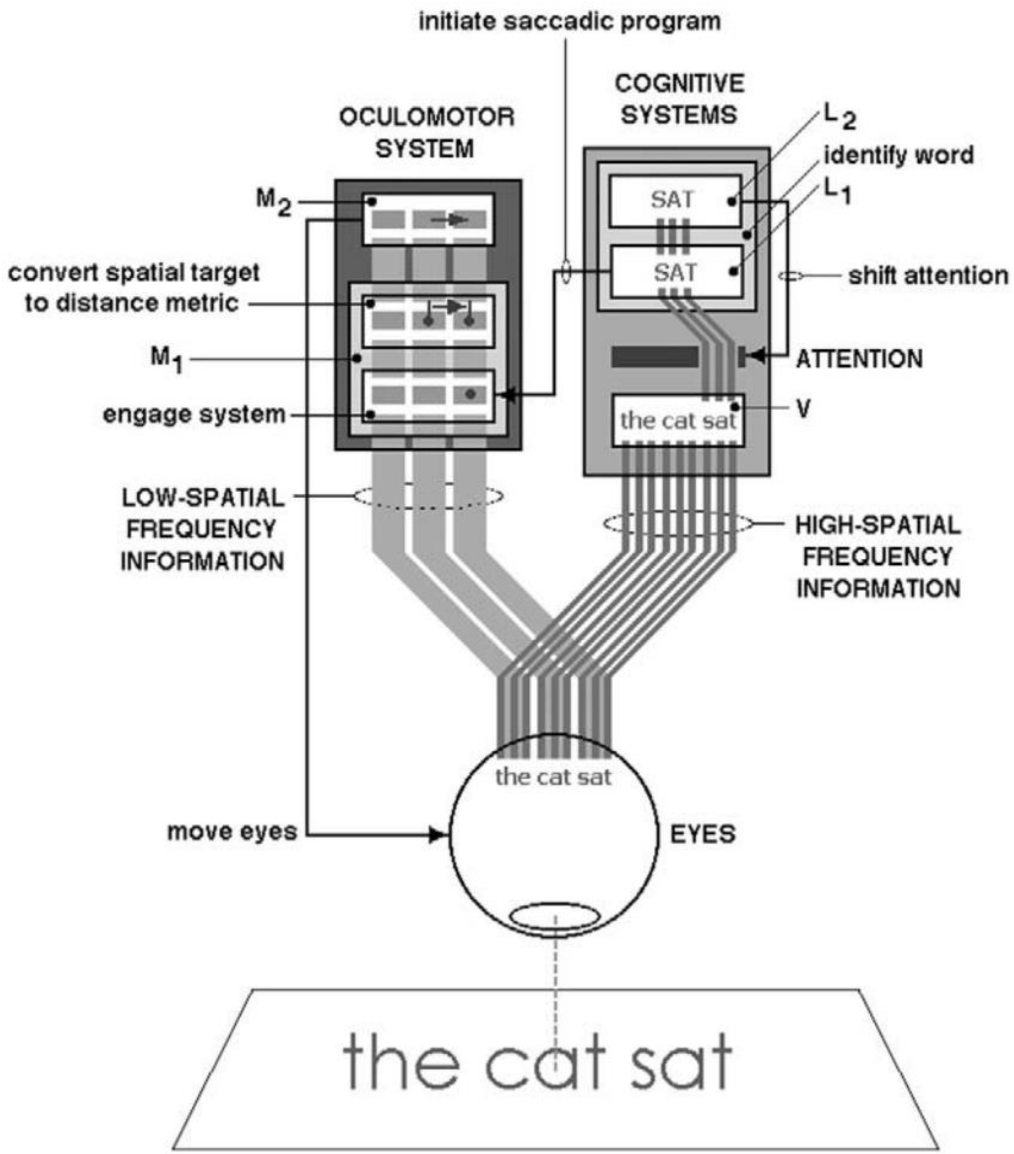
NLP for Modeling Eye Movements in Reading



	Examples	**Output**
Cognitive	E-Z Reader Reichle et al. (1998, 2009)	Fixation
	SWIFT Engbert et al. (2005)	Fixation, transition probability
	SEAM Rabe et al. (2024)	Fixation, transition probability
	OB1-Reader Snell et al. (2018)	Fixation, transition probability
NLP/ML-Based	NEAT Hahn and Keller (2023)	Fixation
	Eyettention Deng, Reich et al. (2023)	Transition probability
	ScanDL Bolliger et al. (2023, 2025)	Fixation
	SP-EyeGan Prasse, Reich et al. (2023)	Raw samples

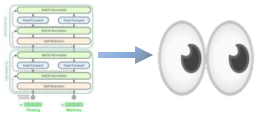
NLP for Modeling Eye Movements in Reading

- E-Z Reader

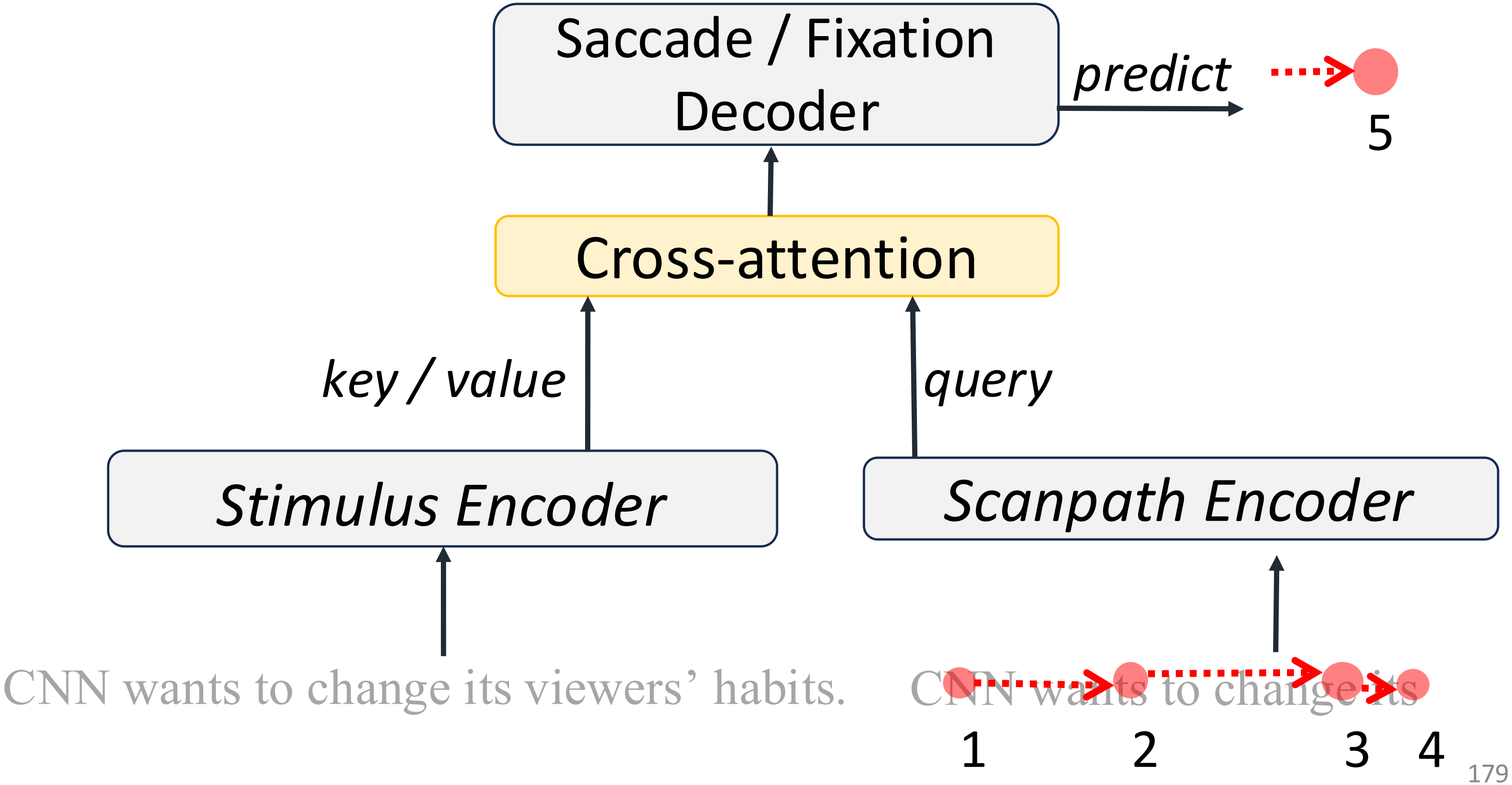


[Pollatsek et al. (2006)]

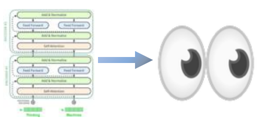
NLP for Modeling Eye Movements in Reading



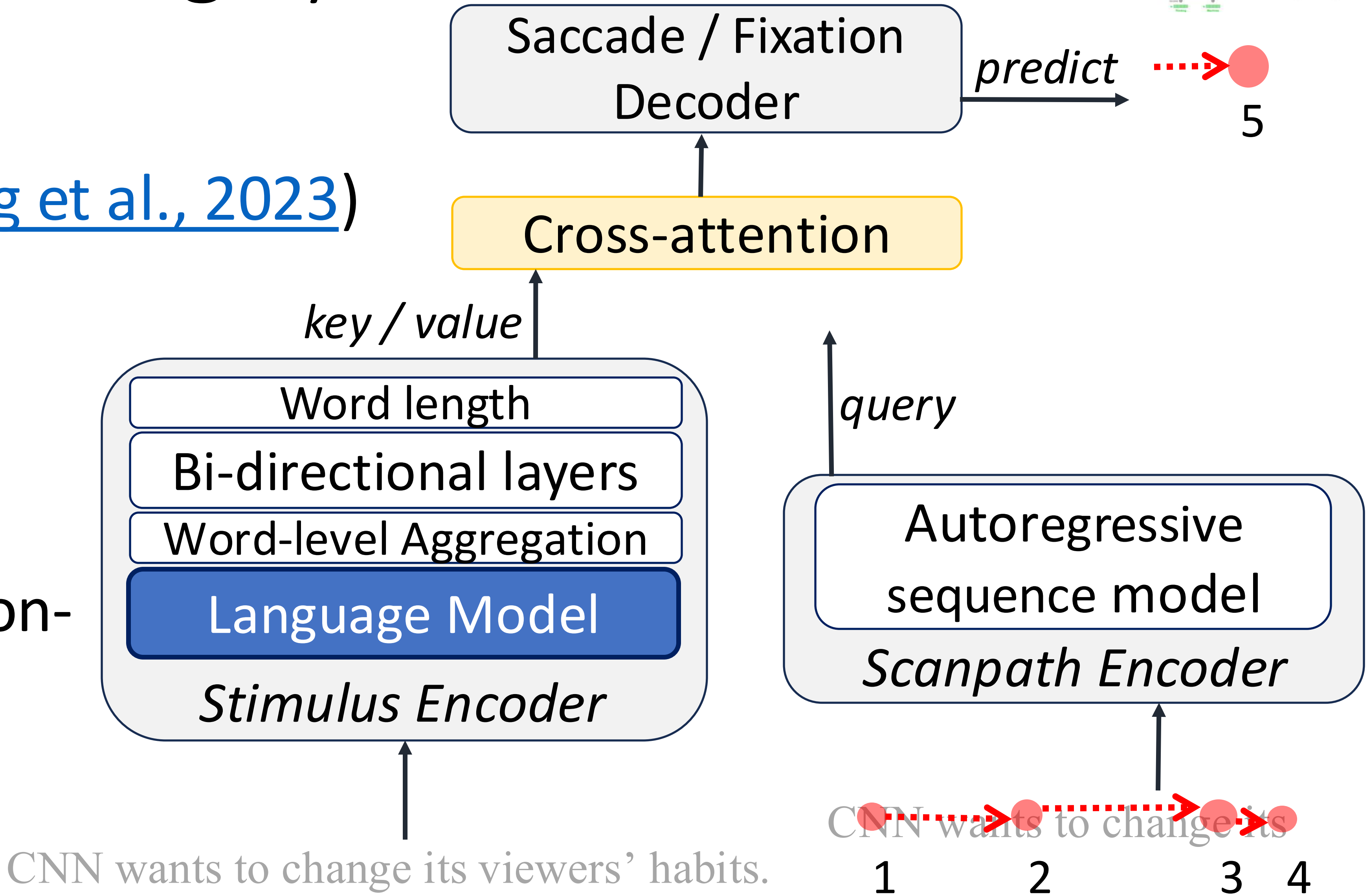
Eyettention (Deng et al., 2023)



NLP for Modeling Eye Movements in Reading



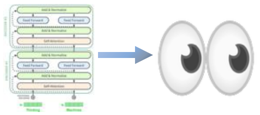
Eyettention ([Deng et al., 2023](#))



Extensions:

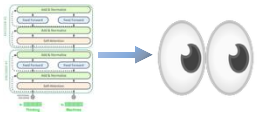
Reader- and population-specific models

NLP for Modeling Eye Movements in Reading



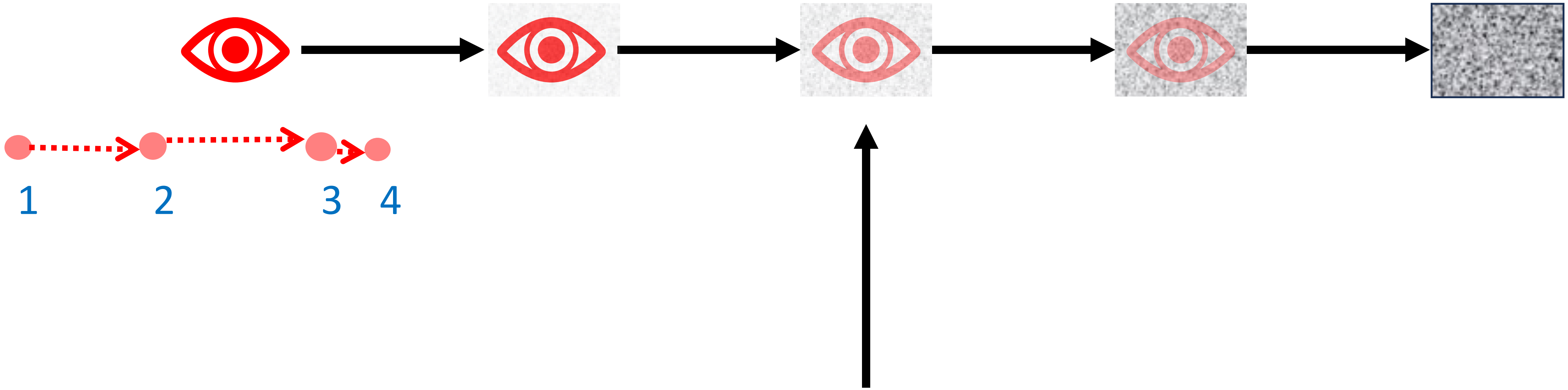
ScanDL, ScanDL 2.0 (Bolliger et al., 2023 ,2025)

NLP for Modeling Eye Movements in Reading



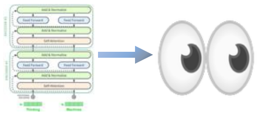
ScanDL, ScanDL 2.0 ([Bolliger et al., 2023](), [2025]())

- How? Discrete input into continuous space



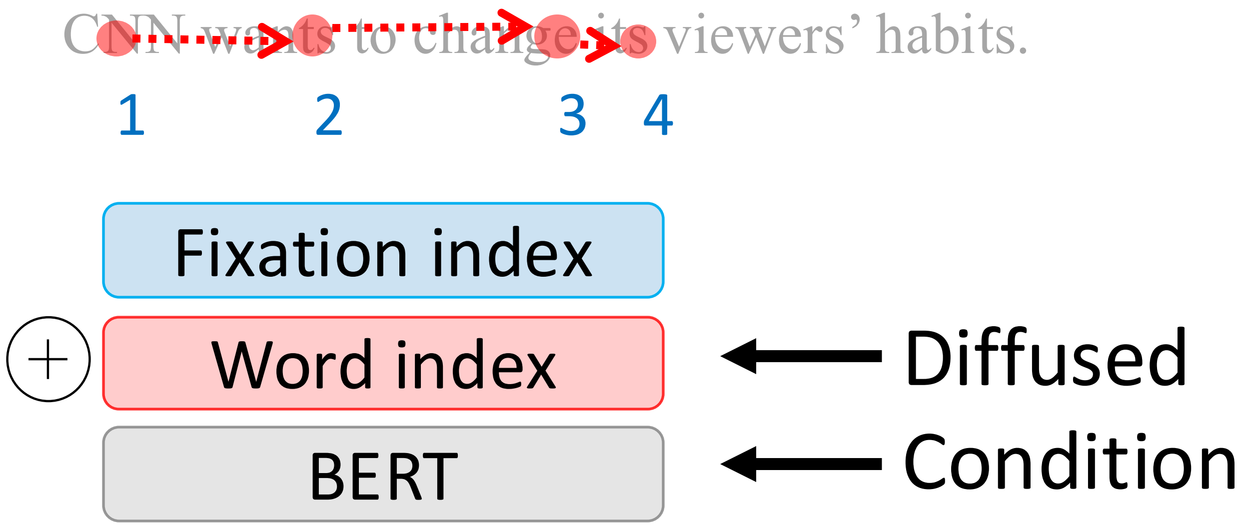
CNN wants to change its viewers' habits.

NLP for Modeling Eye Movements in Reading

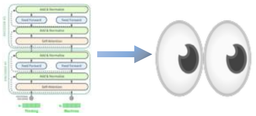


ScanDL, ScanDL 2.0 ([Bolliger et al., 2023](#) ,[2025](#))

- How? Discrete input into continuous space

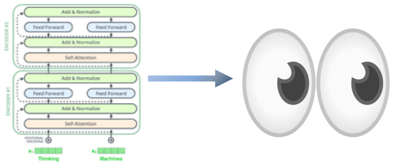


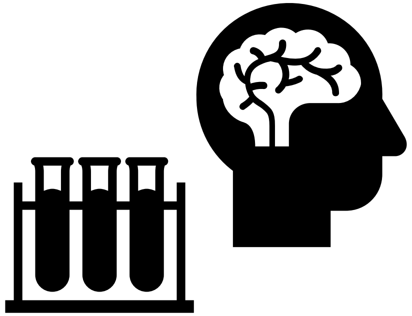
Discussion: NLP for Modeling Eye Movements in Reading



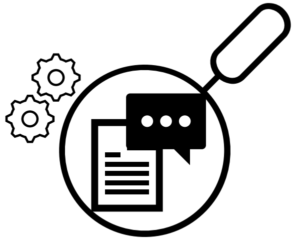
- NLP/ML based models outperform cognitive models
  - But metrics for scanpath generation – nontrivial!
- Making NLP/ML models more cognitively plausible and interpretable

Uses of NLP in Modeling Eye Movements and Human Language Processing



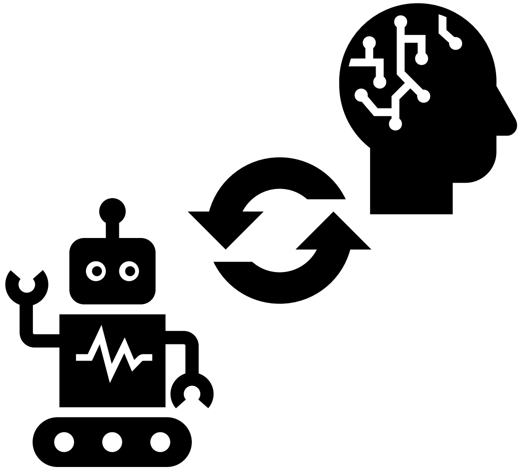


Testing Psycholinguistic Theories



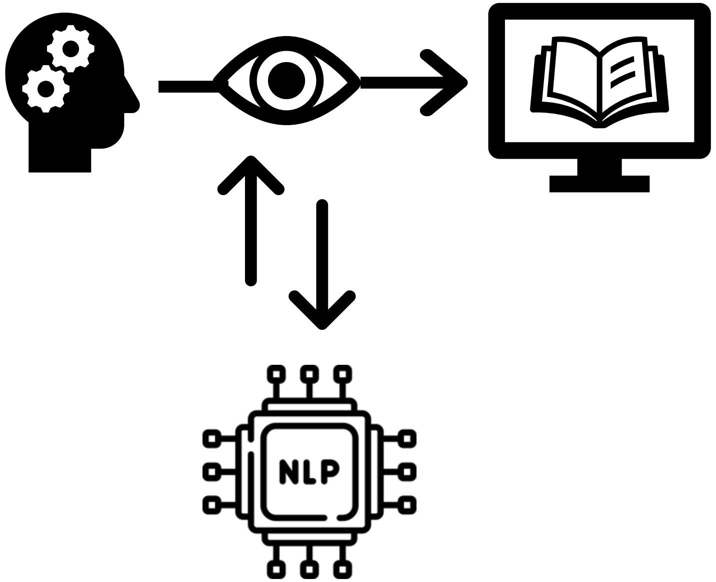
Representations

Linguistic quantities

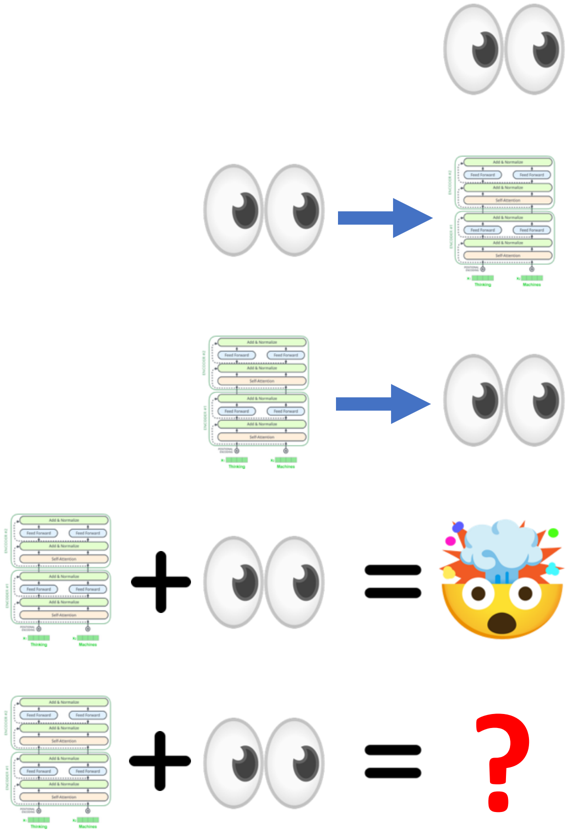


Testing LLM alignment with human language processing

NLP for modeling eye movements in reading

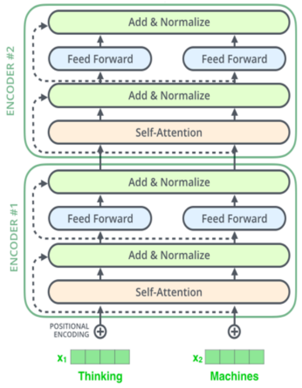


Tutorial Outline



1. Introduction to eye tracking
2. Uses of eye tracking in NLP
3. NLP for eye movement and cognitive modeling
4. **New human centered applications**
5. Outlook and future directions

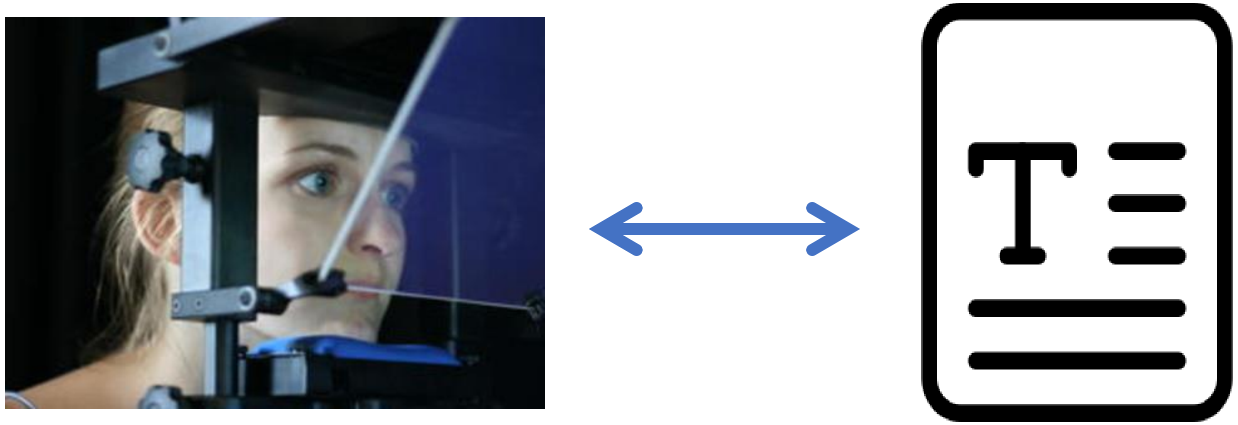
New Human Centered Applications



+ 👀 = 🤯

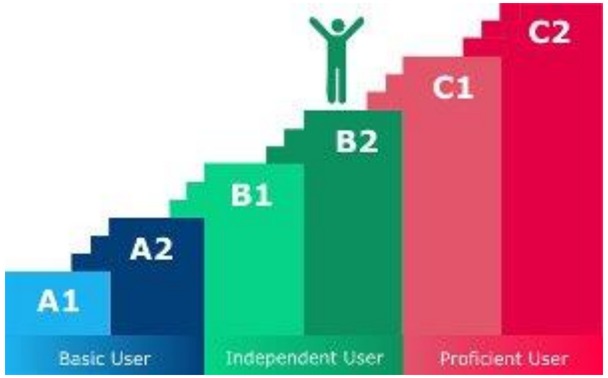
Human Centered NLP with Eye Movements

- Rethinking the future of NLP
- Enabling a wide range of new human centered tasks
- Real-time predictions about reader and their interactions with the text



Human Centered NLP with Eye Movements

Language assessment

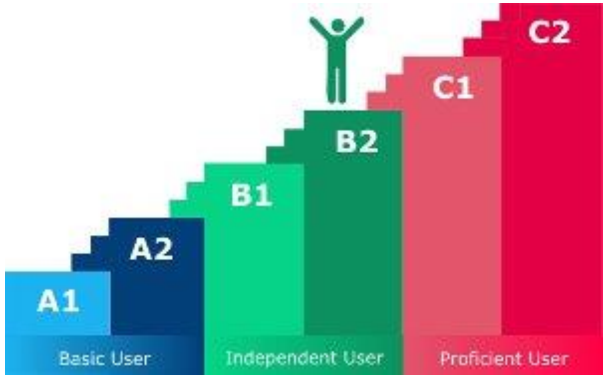


Reading impairment screening and monitoring

Assessment of Reading Comprehension

Human Centered NLP with Eye Movements

Language assessment



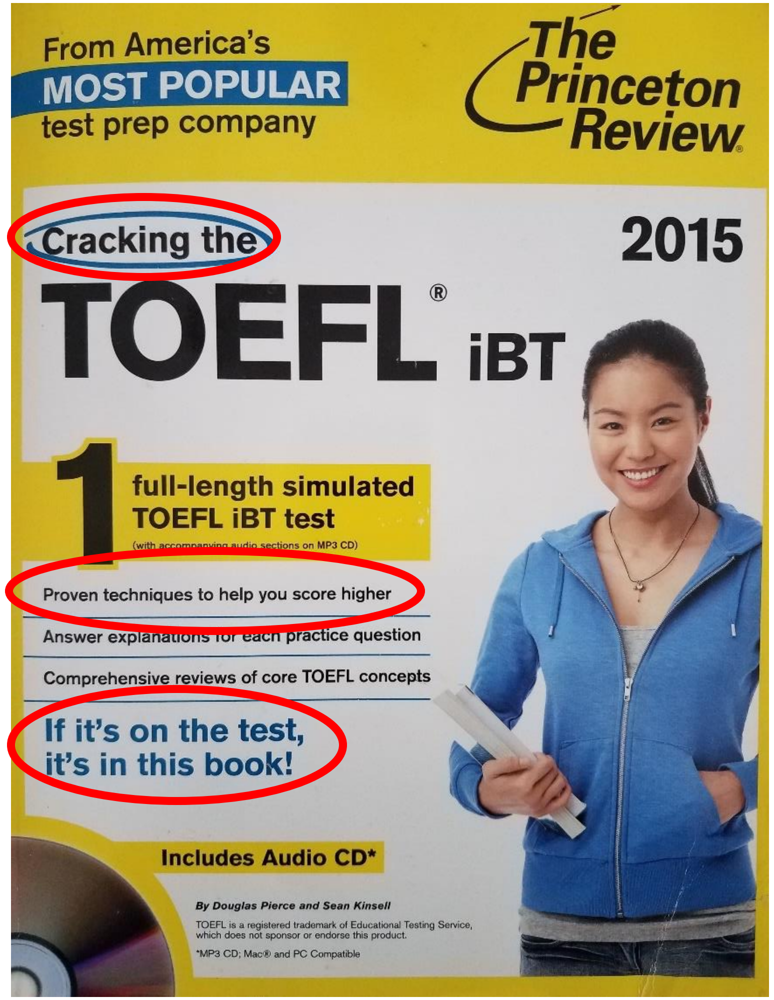
Reading impairment screening and monitoring

Assessment of Reading Comprehension

Language Proficiency Assessment

- Over 2 billion English learners worldwide
- Grammar & vocabulary quizzes
- Reading comprehension
- Listening comprehension
- Essay writing
- …

Language Proficiency Assessment



Language Proficiency Assessment 

- Expensive
- Require test specific preparation
- Cheating
- Manually crafted ad-hoc tasks
- No ability to track language processing online

Language Proficiency Assessment

- Eye movements are informative of L2 language proficiency

Assessing Language Proficiency from Eye Movements in Reading

Yevgeni Berzak
MIT BCS
berzak@mit.edu

Boris Katz
MIT CSAIL
boris@mit.edu

Roger Levy
MIT BCS
rplevy@mit.edu

[Berzak et al. (2018)]

Journal of Optical Technology Vol. 89, Issue 8, pp. 484-489 (2022) • https://doi.org/10.1364/JOT.89.000484

OPTICA PUBLISHING GROUP

Evaluation of level of foreign language proficiency based on eye movement data

V. A. Demareva, A. V. Golubinskaya, Yu. A. Edeleva, and R. V. Golubin

[Demareva et al. (2022)]

Inferring Search User Language Proficiency from Eye Gaze Data

Ben Steichen
California State Polytechnic University, Pomona
bsteichen@cpp.edu

Wilsen Kosasih
California State Polytechnic University, Pomona
wkosasih@cpp.edu

Christian Becerra
California State Polytechnic University, Pomona
ceb@cpp.edu

[Steichen et al. (2024)]

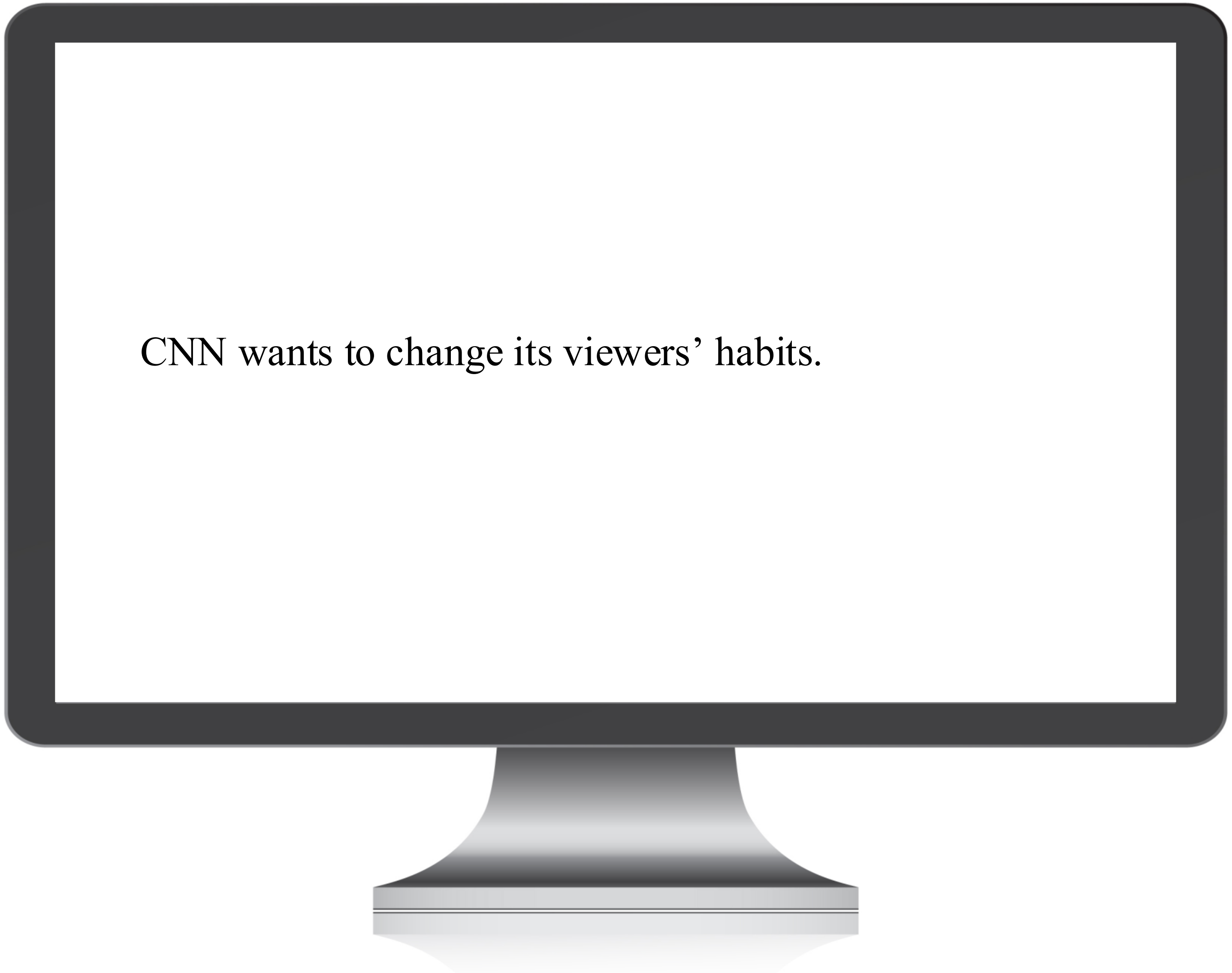
Predicting First-Language and Second-Language Proficiency Using Eye Fixation Data and Demographic Information: Assumptions, Data Representations, and Methods

Publisher: IEEE Cite This PDF

Soroosh Shalileh ; Matvey Kairov ; Ranga Baminiwatte ; Olga Parshina ; Olga Dragoy All Authors

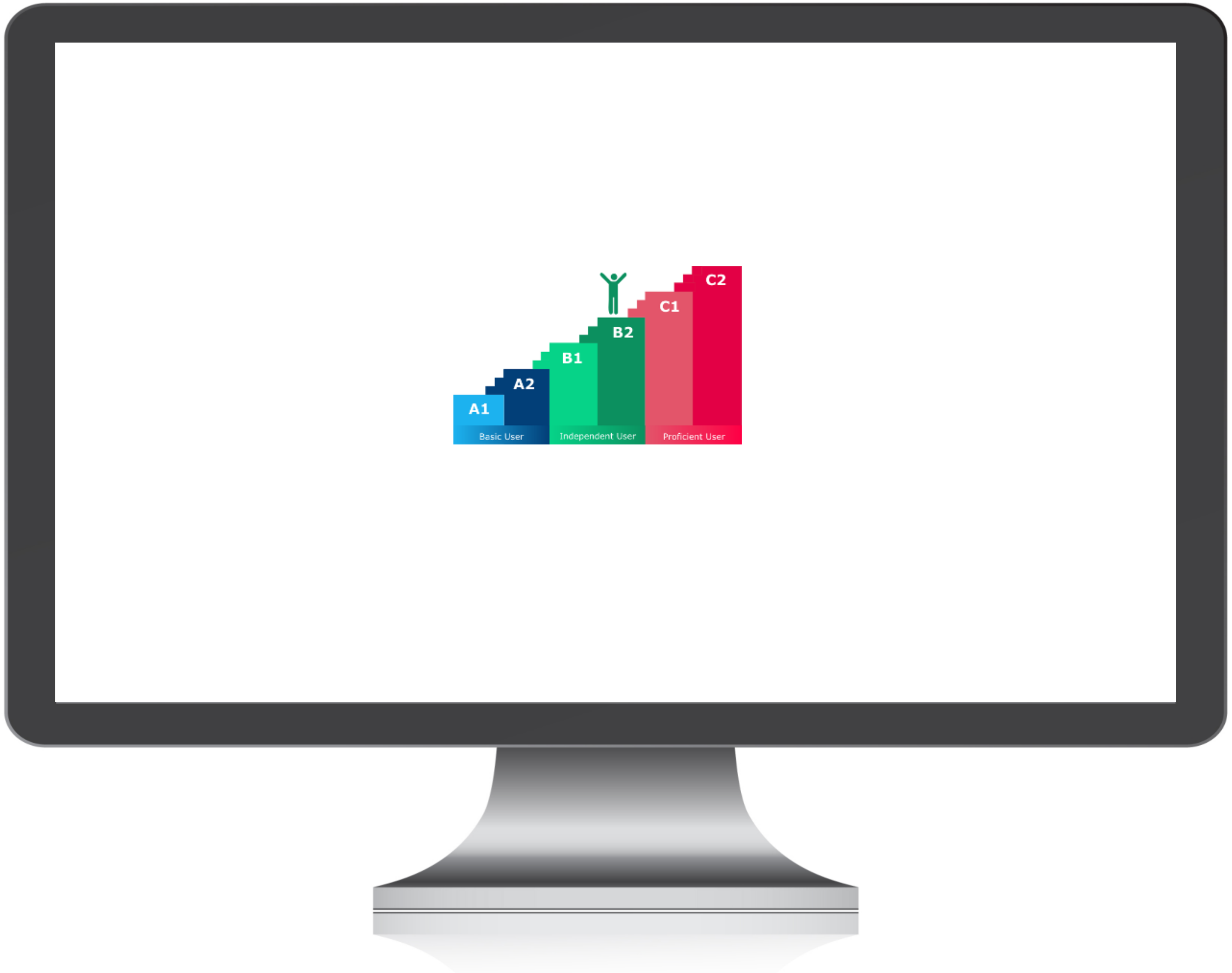
[Shalileh et al. (2024)]

Our Test



Our Test

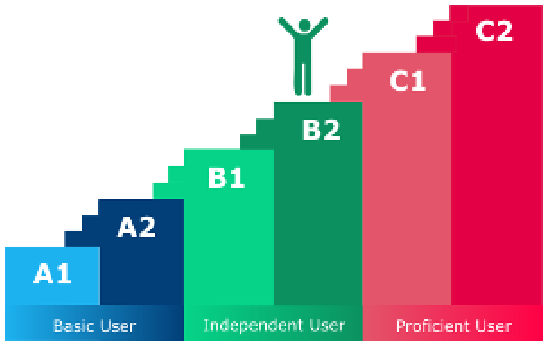
Our Test



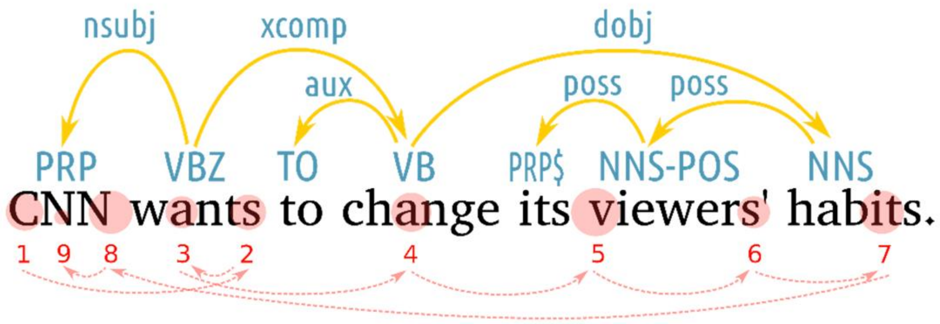
A1
A2
B1
B2
C1
C2
Basic User
Independent User
Proficient User

Our Test

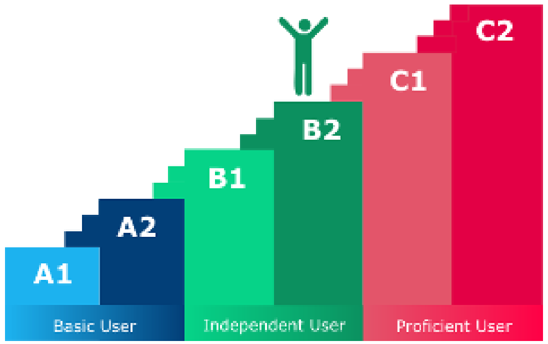
Ordinary Reading



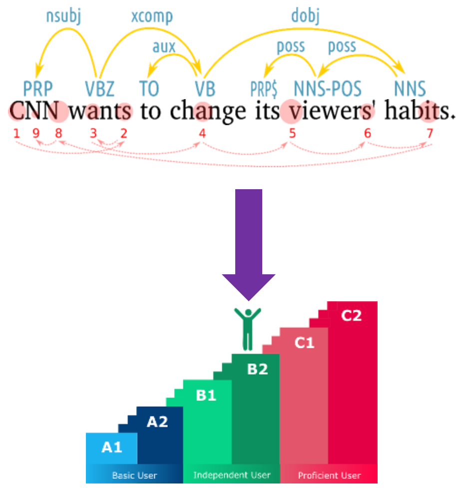
Our Test



Linguistic Processing

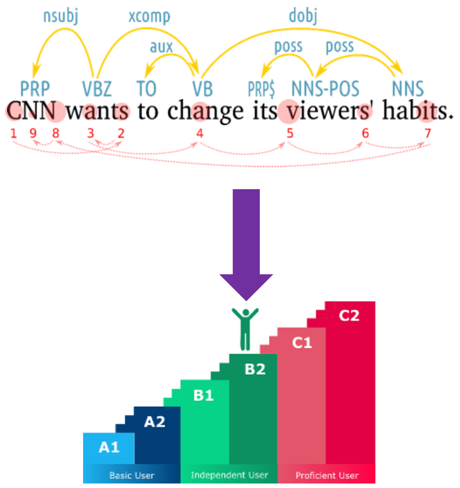


Our Test



- ~~Expensive~~
- ~~Require test specific preparation~~
- ~~Cheating~~
- ~~Manually crafted ad-hoc tasks~~
- Ability to track language processing online

Our Test

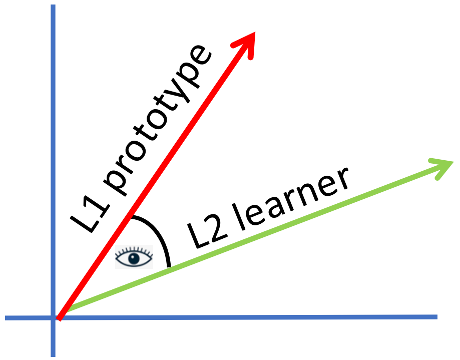


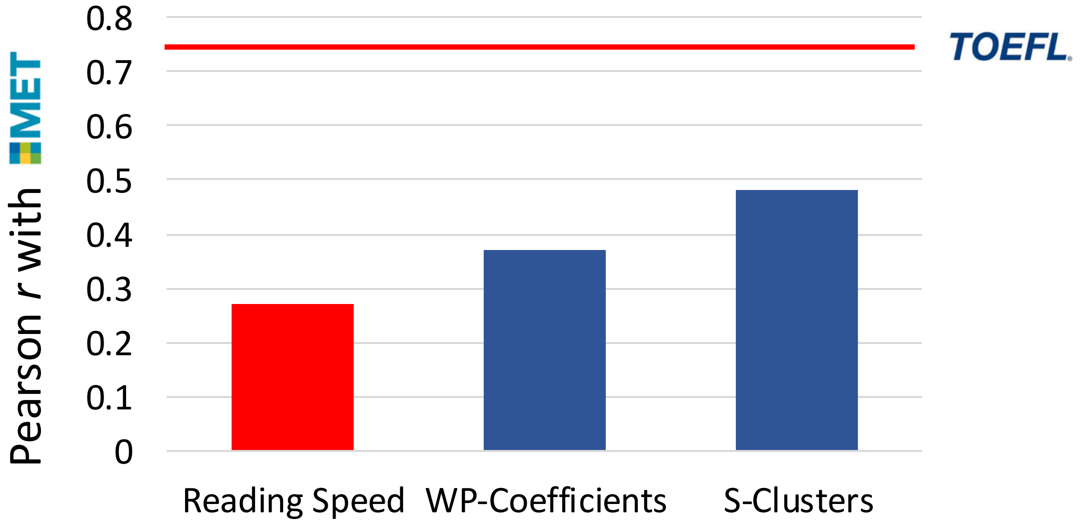
ESL proficiency ≈ Similarity of reading patterns to native speakers of English

EyeScore

[Berzak et al. (2018)](#)

- Extract eye movement features for each participant
- Compute English L1 “prototype”
- **EyeScore** = cosine similarity to L1 prototype





But

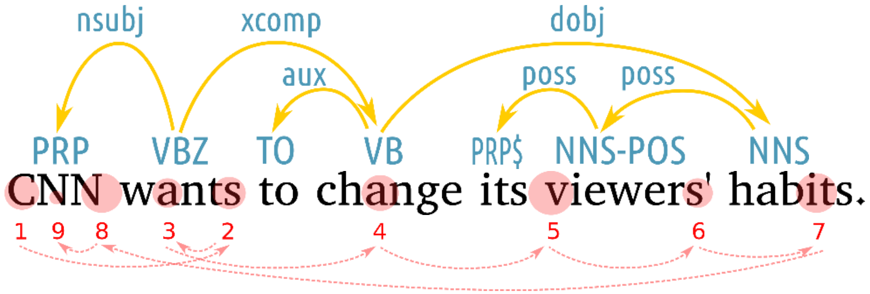
I

MET

Predicting Scores on Standardized Tests

Berzak et al. (2018)

- Eye movement in reading can be used to predict scores of specific external proficiency tests



MET

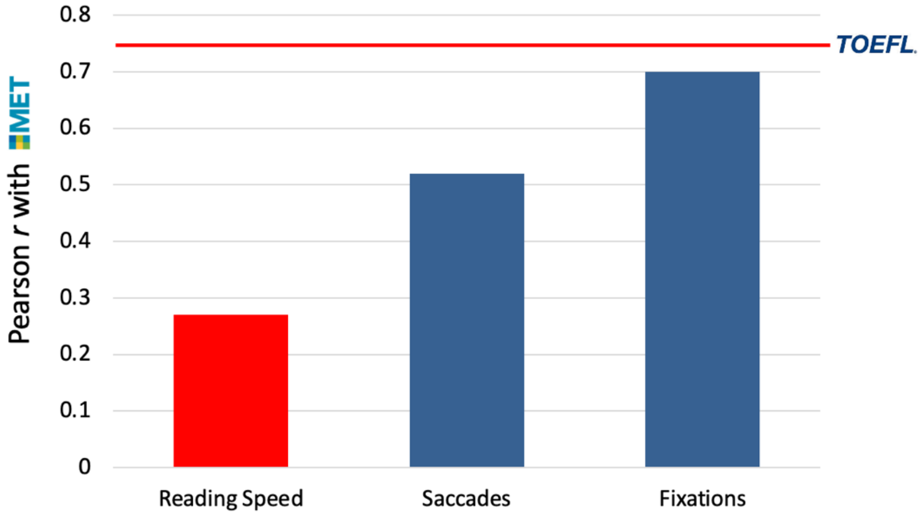
ETS TOEFL

Predicting Scores on Standardized Tests

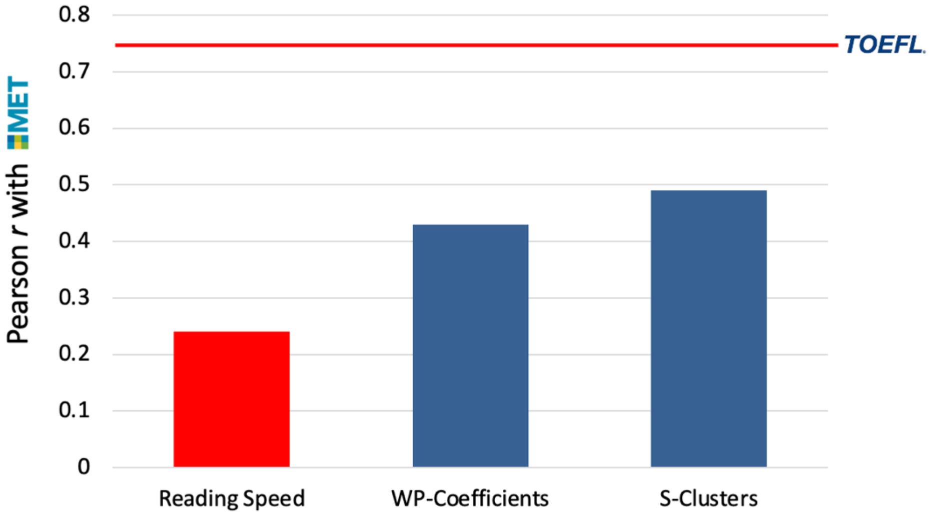
[Berzak et al. (2018)](#)

- Eye movement in reading can be used to predict scores of specific external proficiency tests

Seen Texts



New Texts



Human Centered NLP with Eye Movements

Language assessment

Reading impairment Screening and monitoring

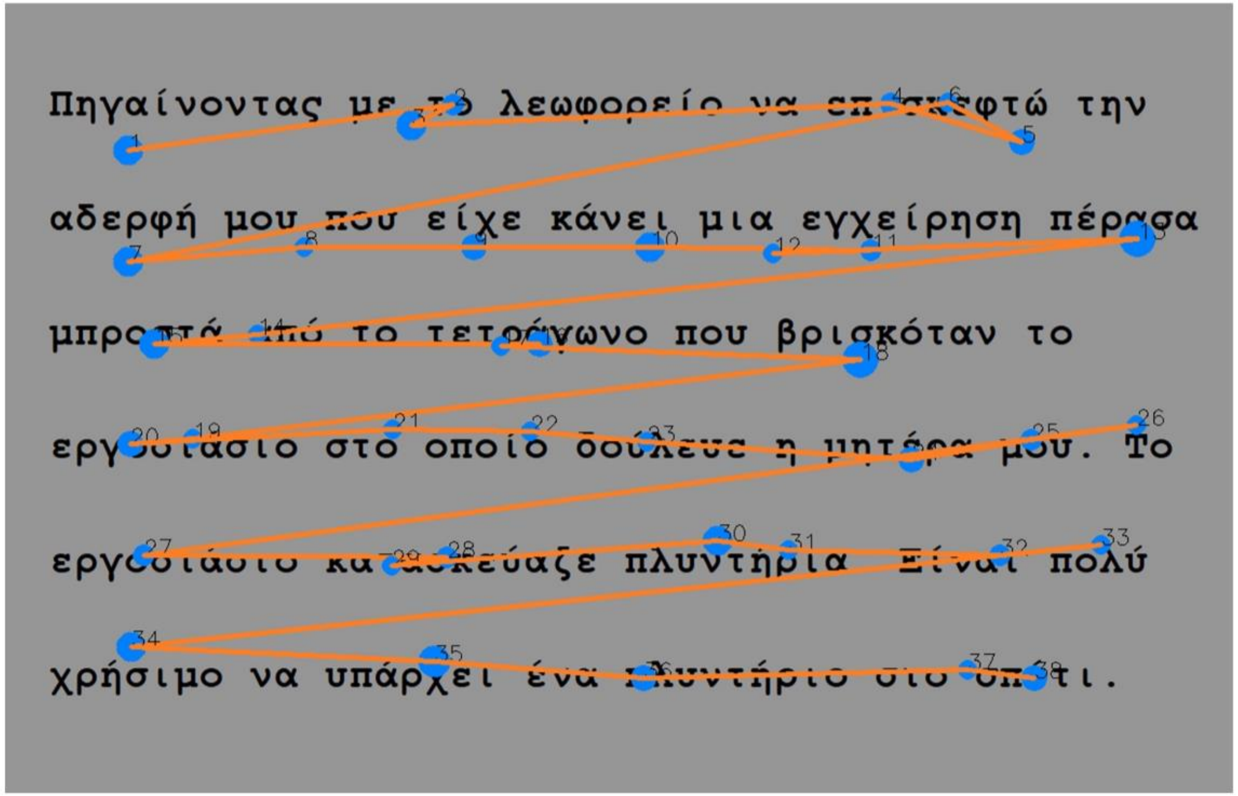
Assessment of Reading Comprehension

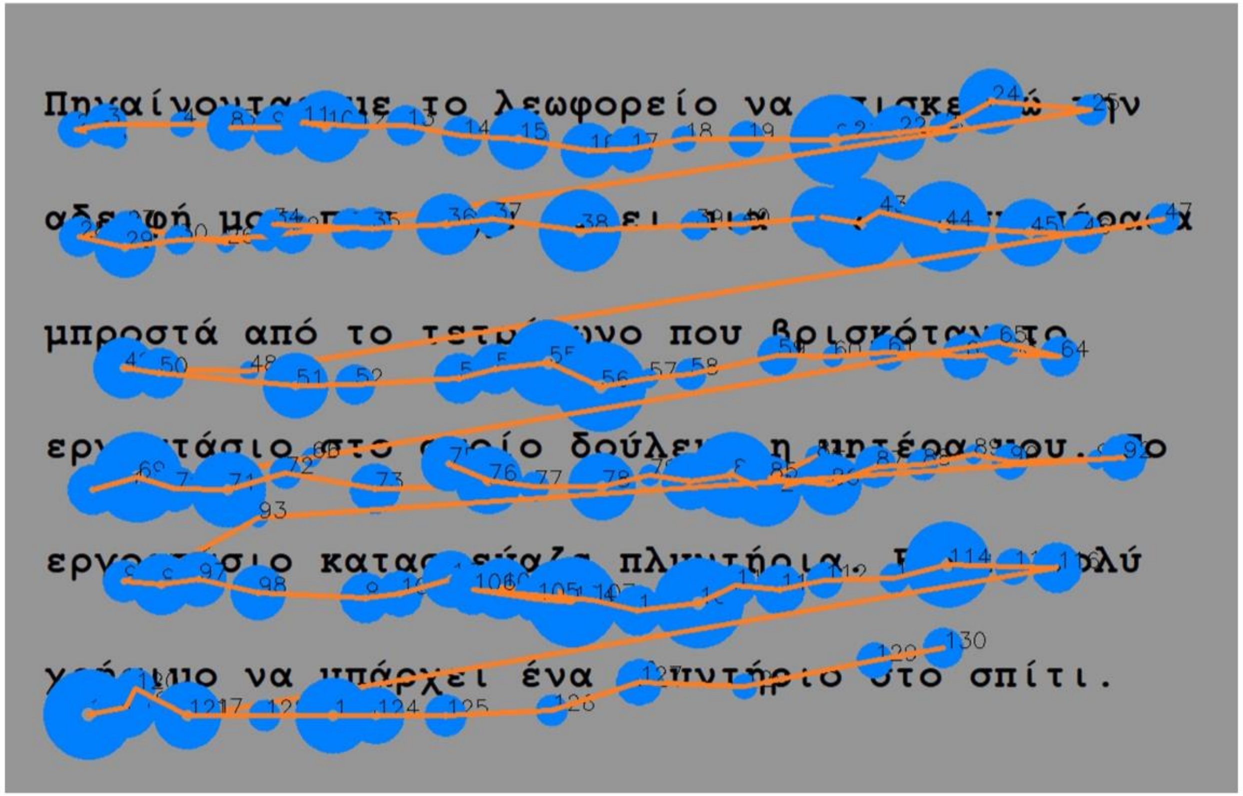
Reading Impairments Developmental Dyslexia

"... *impairment in reading [which] is characterised by significant and persistent difficulties in learning academic skills related to reading, … [and] is not due to a disorder of intellectual development, sensory impairment (vision or hearing), … .*" (WHO, 2025)

- Affects approx. 7-10% of the population (Catts et al., 2005, Peterson & Pennington, 2012; Moll et al., 2014)
- Early detection and intervention is key (Snowling, 2012; Torgesen, 2002)
- Existing testing batteries must be administered by a trained specialist.

Reading Impairments
Developmental Dyslexia





[Smyrnakis et al. (2017)](#)

Reading Impairments
Developmental Dyslexia

Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures

Luz Rello
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
luzrello@cs.cmu.edu

Miguel Ballesteros
Natural Language Processing Group
Universitat Pompeu Fabra
miguel.ballesteros@upf.edu

Rello and Ballesteros (2014)

Screening for Dyslexia Using Eye Tracking during Reading

Mattias Nilsson Benfatto, Gustaf Öqvist Seimyr, Jan Ygge, Tony Pansell, Agneta Rydberg, Christer Jacobson

Published: December 9, 2016 • https://doi.org/10.1371/journal.pone.0165508

Nilsson Benfatto et al. (2016)

Predictive Model for Dyslexia from Fixations and Saccadic Eye Movement Events

A Jothi Prabha, R Bhargavi

Jothi Prabha and Bhargava (2020)

Eye-tracking based classification of Mandarin Chinese readers with and without dyslexia using neural sequence models

Patrick Haller[1], Andreas Säuberli[1], Sarah E. Kiener[1]
Jinger Pan[3], Ming Yan[4], Lena A. Jäger[1,2]

Haller et al. (2022)

Dyslexia Prediction from Natural Reading of Danish Texts

Marina Björnsdóttir, Nora Hollenstein, Maria Barrett

Björnsdóttir et al. (2023)

Identifying dyslexia in school pupils from eye movement and demographic data using artificial intelligence

Soroosh Shalileh[1]*, Dmitry Ignatov[2], Anastasiya Lopukhina[3], Olga Dragoy[1,4]

Shalileh et al. (2023)

Automatic detection of dyslexia based on eye movements during reading in Russian

Anna Laurinavichyute Anastasiya Lopukhina David R. Reich

Laurinavichyute et al. (2025)

Reading Impairments
Developmental Dyslexia Datasets

[Björnsdóttir et al. (2023)](#)

- 58 L1, L2 or adults with dyslexia reading paragraphs

[Shalileh et al. (2023)](#)

- 300+ children with or without dyslexia
- Two different assessments

Reading Impairments
Developmental Dyslexia

- Scale with few eye trackers
  - Cheaper alternatives possible
- Beyond screening

Human Centered NLP with Eye Movements

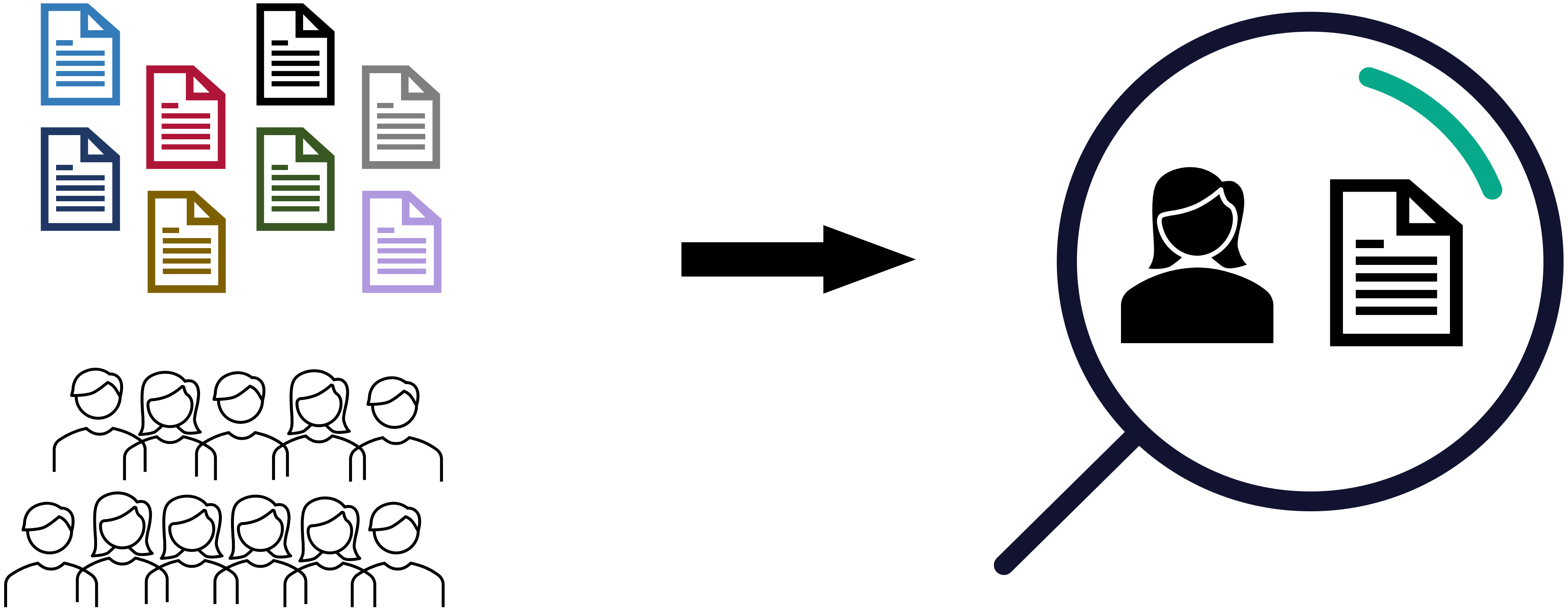
Language assessment

Reading impairment diagnostics and monitoring

Assessment of Reading Comprehension

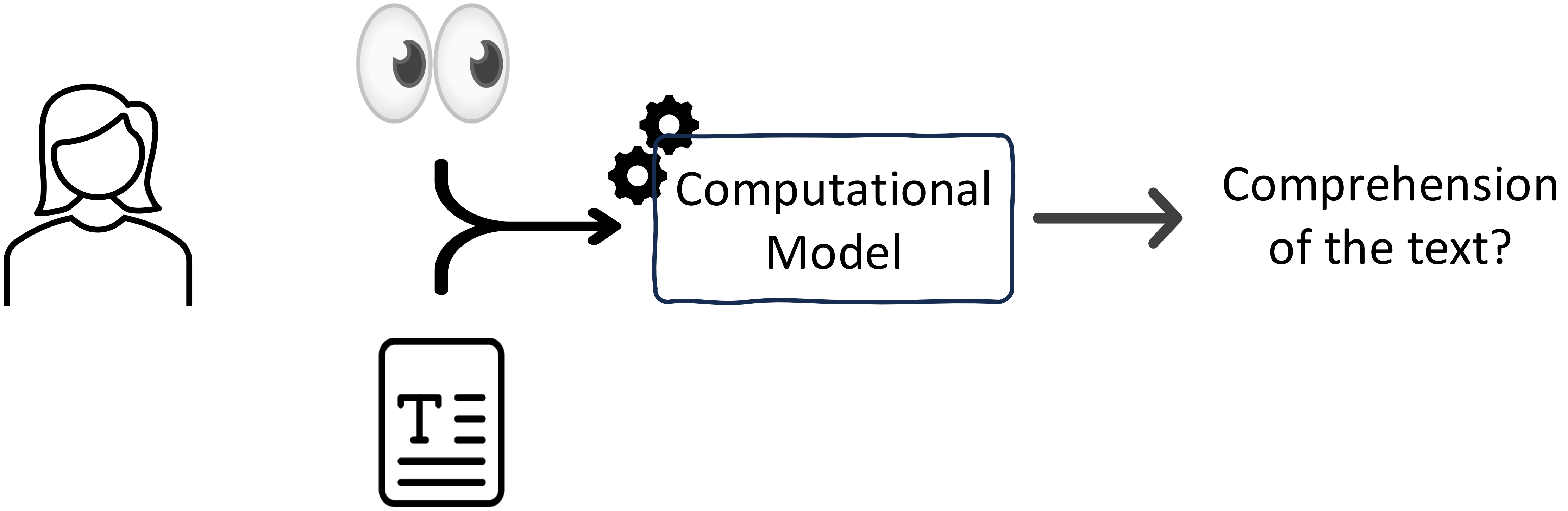
Reading Comprehension

A Holy Grail in psycholinguistics



[Just and Carpenter (1980); Among others]

Reading Comprehension



Reading Comprehension

Towards Predicting Reading Comprehension From Gaze Behavior

Seoyoung Ahn, Stony Brook University, Stony Brook, New York
Conor Kelton, Stony Brook University, Stony Brook, New York
Aruna Balasubramanian, Stony Brook University, Stony Brook, New York
Gregory Zelinsky, Stony Brook University, Stony Brook, New York

Ahn et al. (2020)

Fine-Grained Prediction of Reading Comprehension from Eye Movements

Omer Shubi[1], Yoav Meiri[1], Cfir Avraham Hadar[1], Yevgeni Berzak[1,2]

Shubi et al. (2024)

Above chance performance but highly challenging task!

Inferring Native and Non-Native Human Reading Comprehension and Subjective Text Difficulty from Scanpaths in Reading

Authors: David Robert Reich, Paul Prasse, Chiara Tschirner, Patrick Haller, Frank Goldhammer, Lena A. Jäger
Authors Info & Claims

Reich et al. (2022)

Eye Movements as Images: A Multimodal Framework for Eye Movements Representation

Publisher: IEEE
Cite This
PDF

Dongsen Zhang; Peipei Li; Zekun Li; Yiwei Ru; Huijia Wu; Zhaofeng He
All Authors

Zhang et al. (2025)

Human Centered NLP with Eye Movements Open Frontiers

Information accessibility

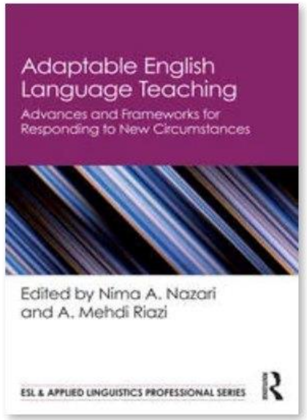
Education and digital learning

Human-machine communication

Interactive online systems

This is happening now!

Eye tracking is starting to gain traction in education research!



Chapter

New Approaches to the Assessment of English as an Additional Language

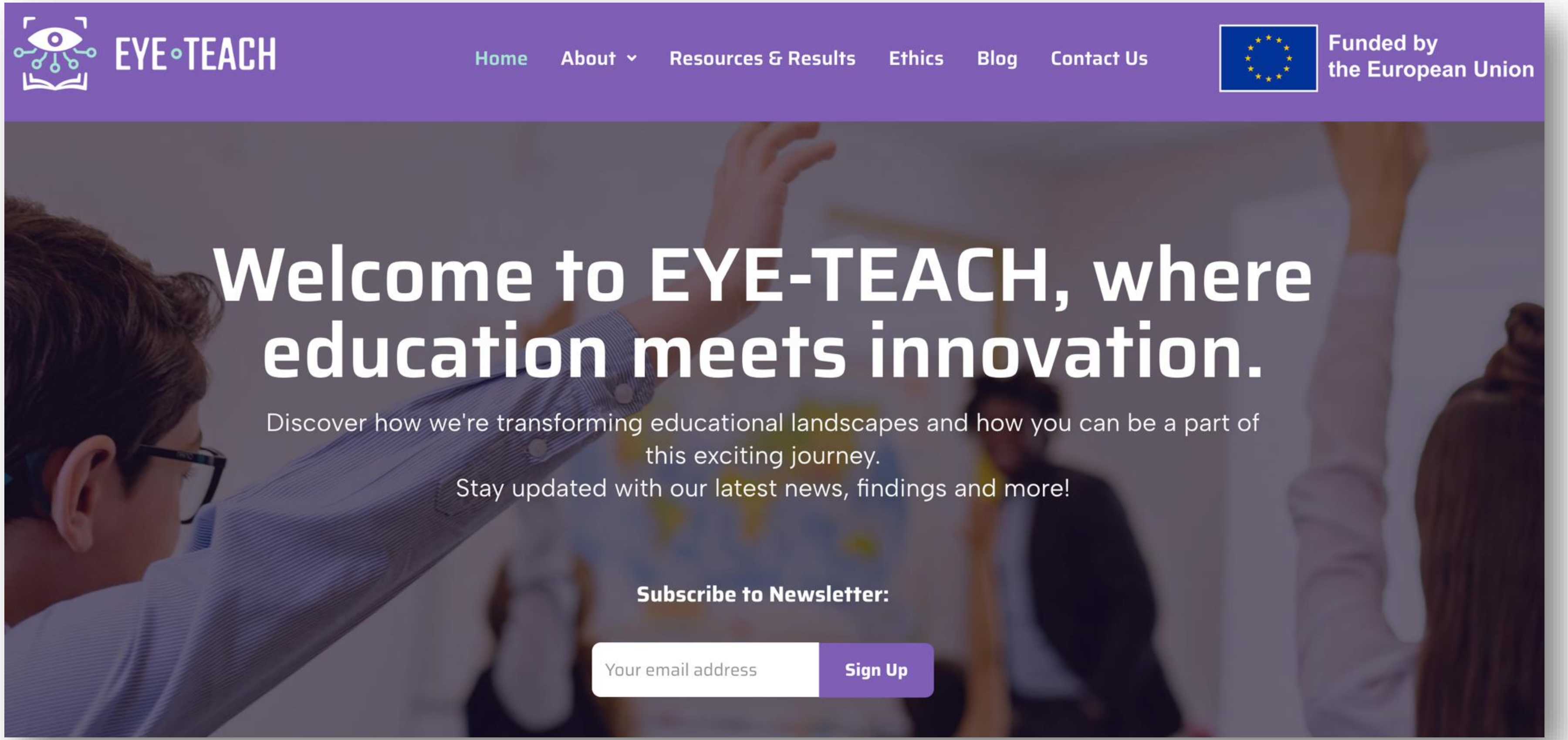
By Graham Seed, Angeliki Salamoura, Nick Saville

Book Adaptable English Language Teaching

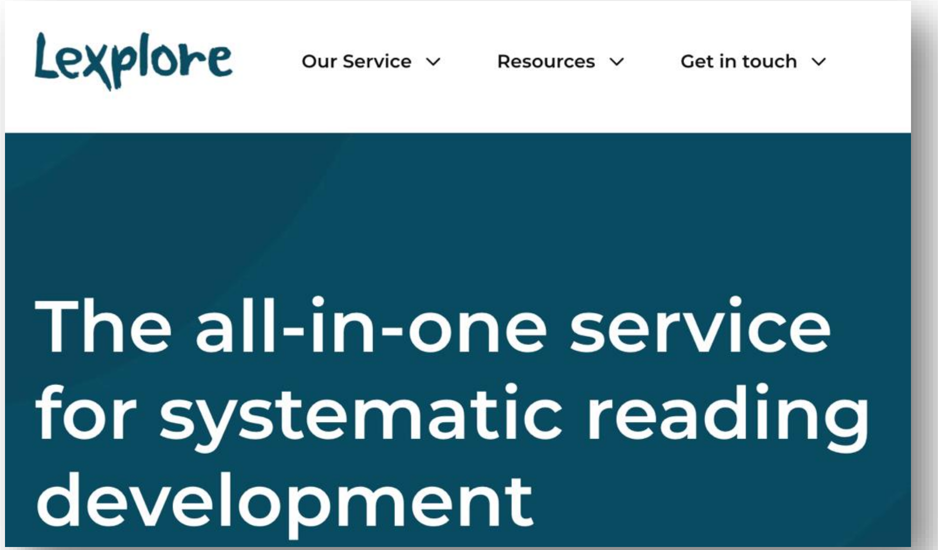
Edition	1st Edition
First Published	2024
Imprint	Routledge
Pages	17
eBook ISBN	9781003361701

Share

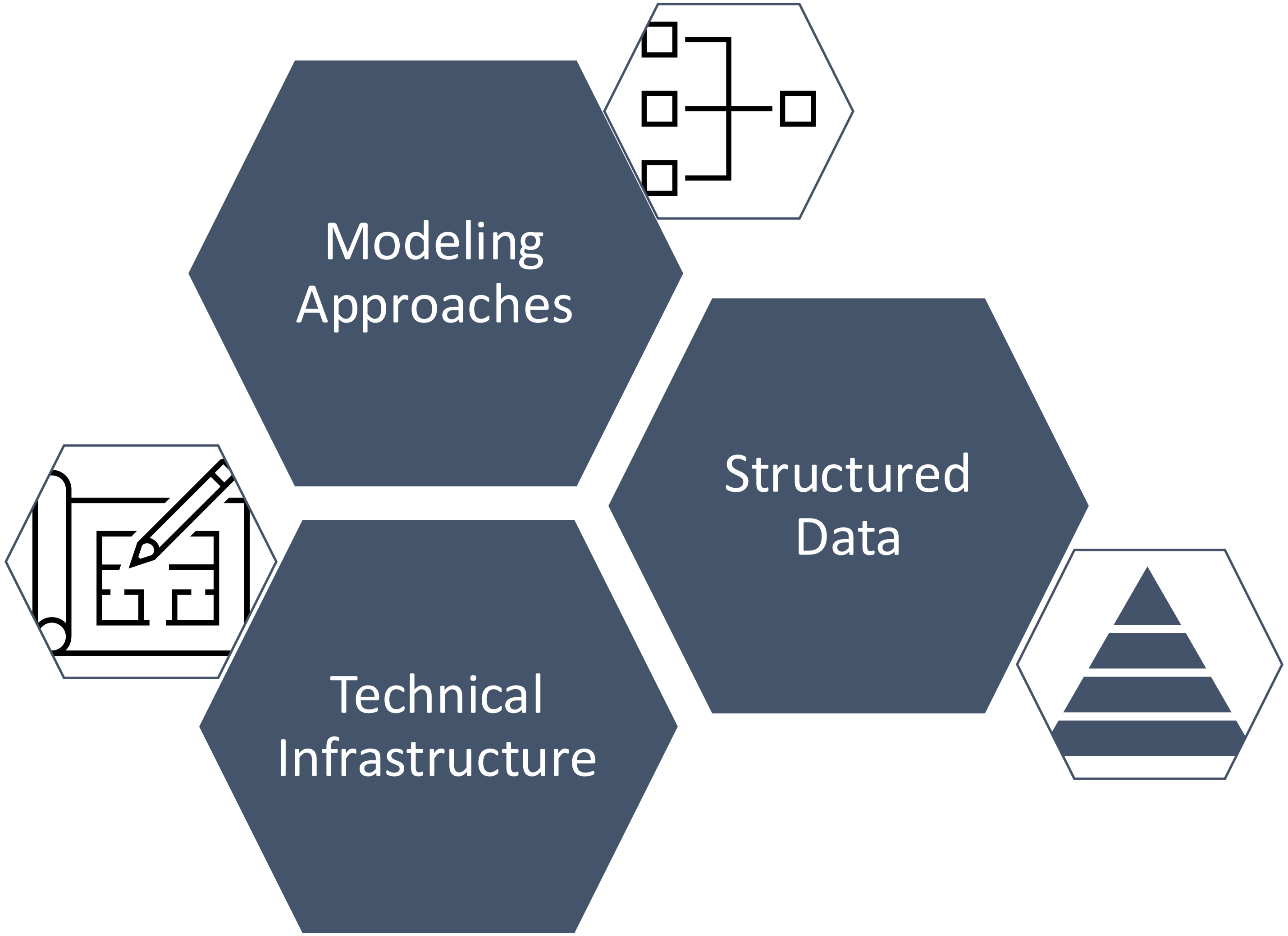
Eye tracking is starting to gain traction in education research!



Eye tracking is starting to gain traction in education applications!



Human Centered NLP with Eye Movements



Relevant to any reader / reader-text task

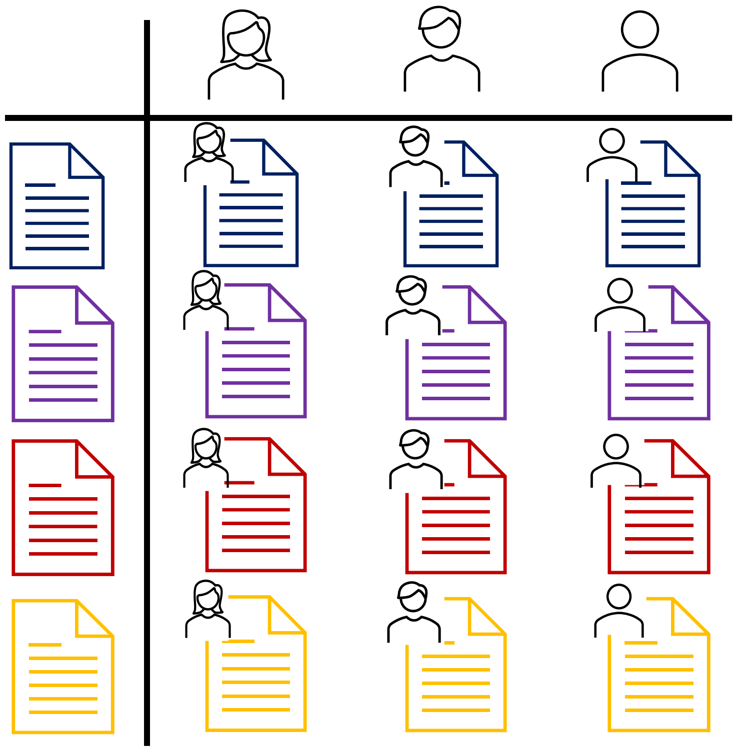
Data is **not** iid – it is **structured**

Implications for:

- **Statistical modeling**
- Training and Evaluations
- Applications

t-test → mixed effects models

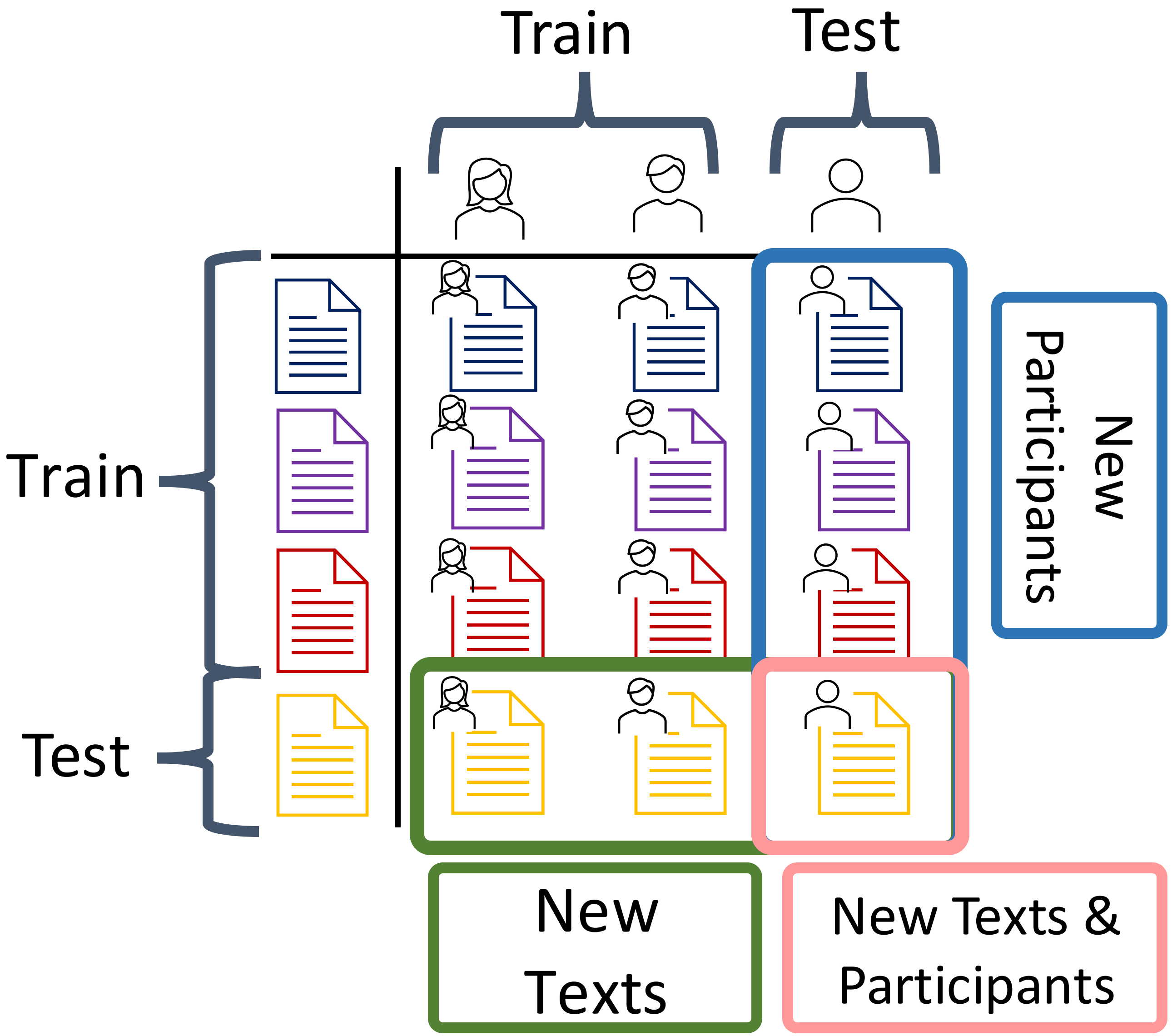
[Shravan Vasishth's Intro Stats course]



Data is **not** iid – it is **structured**

Implications for:

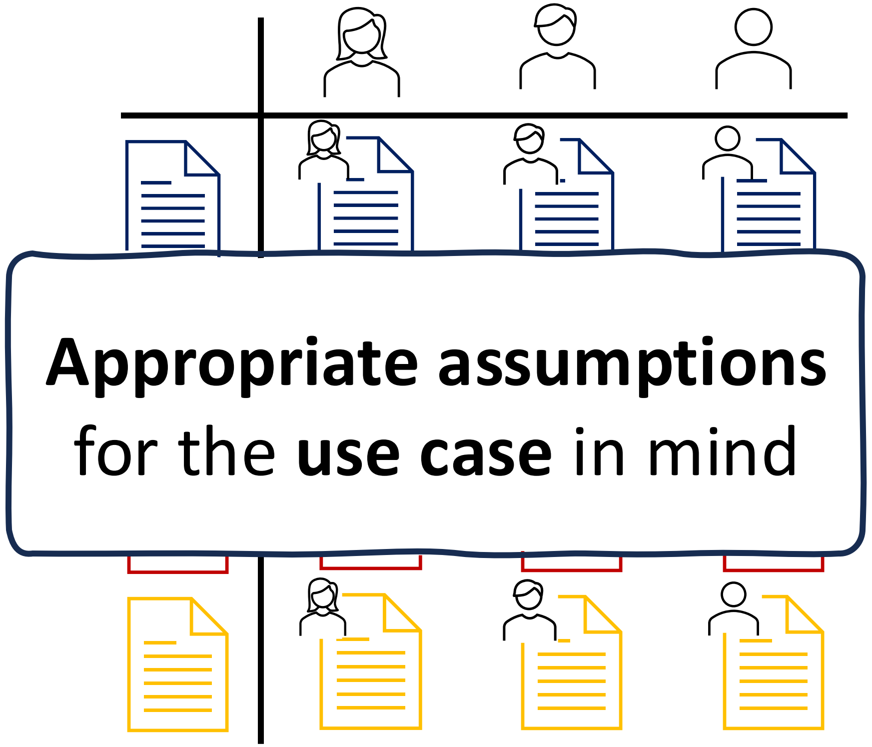
- Statistical modeling
- **Training and Evaluations**
- Applications



Data is **not** iid – it has **structure**

Implications for:

- Statistical modeling
- Training and Evaluations
- **Applications**



Human Centered NLP with Eye Movements Open Frontiers

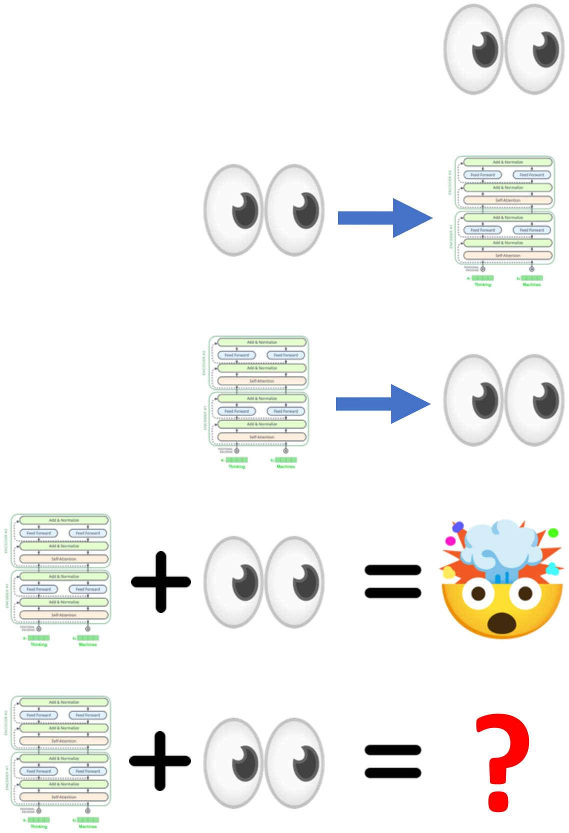
Information accessibility

Education and digital learning

Human-machine communication

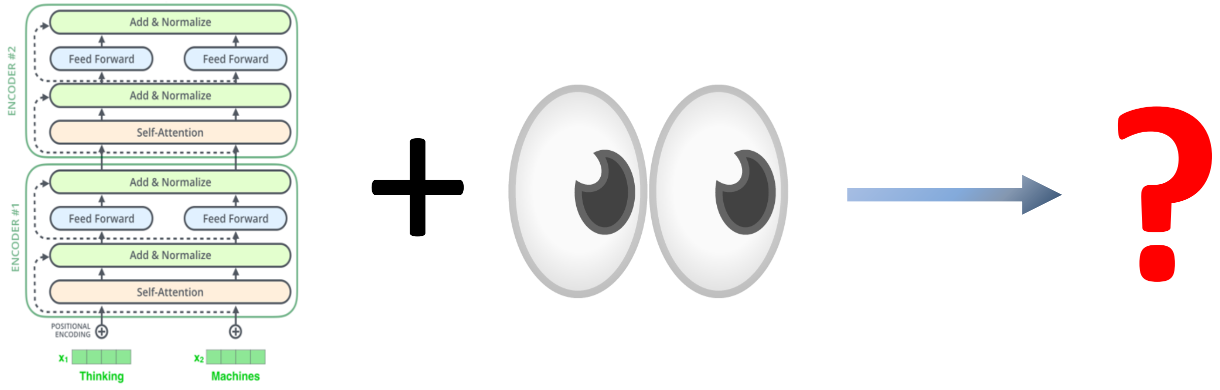
Interactive online systems

Tutorial Outline

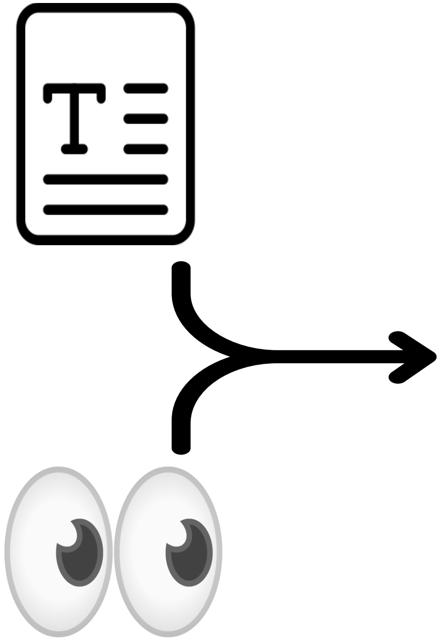


1. Introduction to eye tracking
2. Uses of eye tracking in NLP
3. NLP for eye movement and cognitive modeling
4. New human centered applications
5. Outlook and future directions

Directions for Future Research



The Future: Text + Gaze Multimodal Models



Highly open and challenging problem

Eye movements representation

Alignment with text

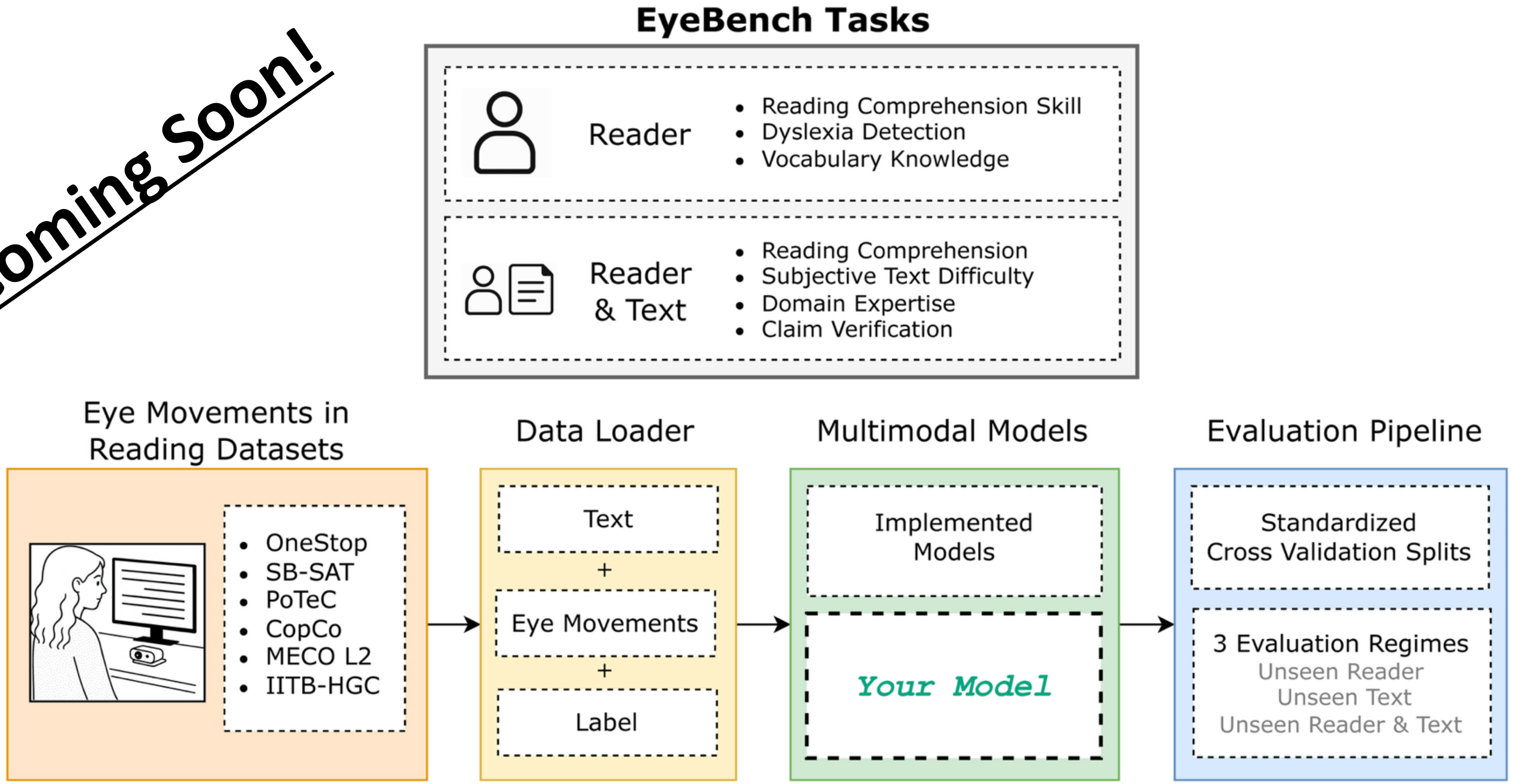
Low resource settings

First-principles cognitive modeling

How to get involved? EyeBench!

EyeBench: Predictive Modeling from Eye Movements in Reading

Coming Soon!

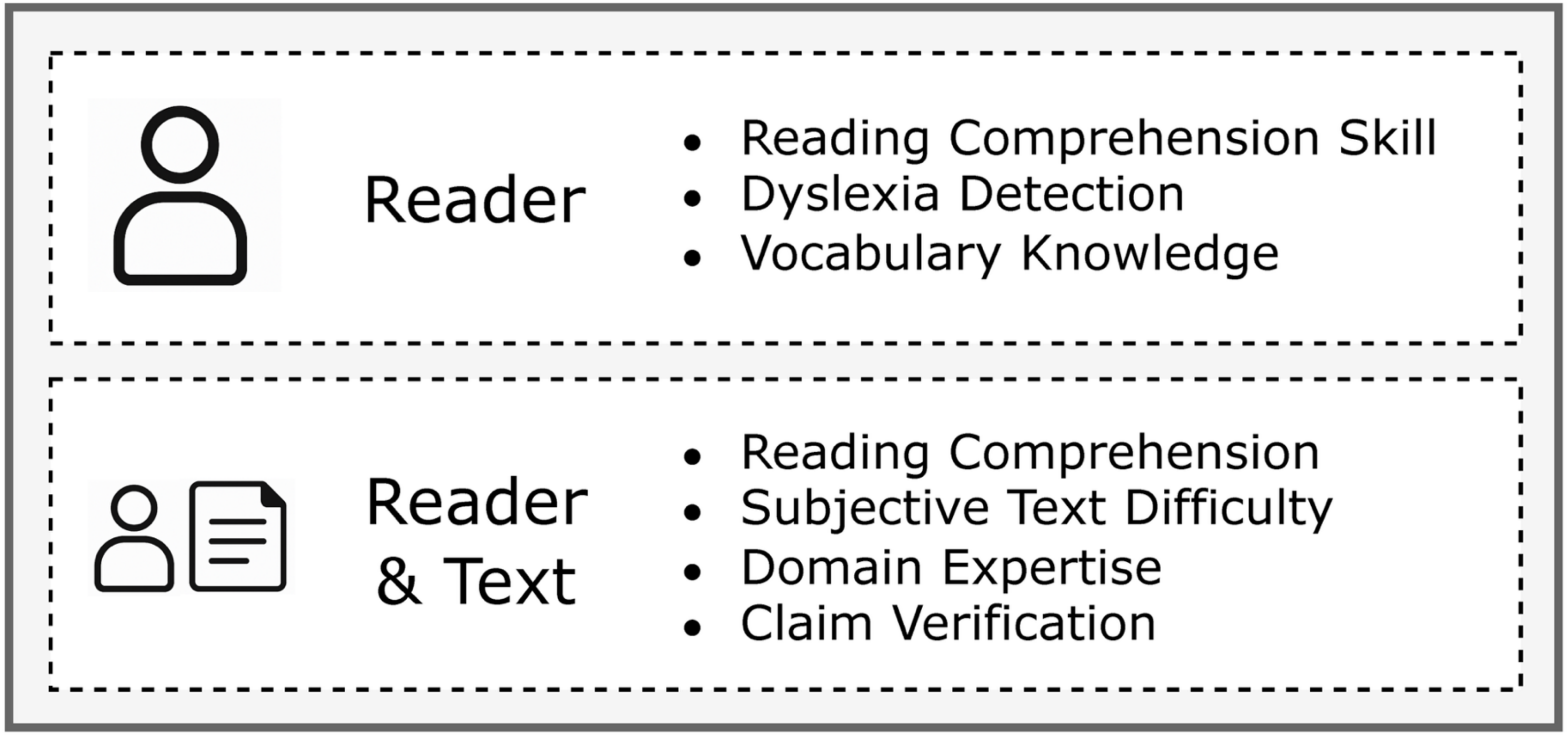


github.com /EyeBench

Shubi, Reich et al. (in prep)

How to get involved? EyeBench!

EyeBench: Predictive Modeling from Eye Movements in Reading



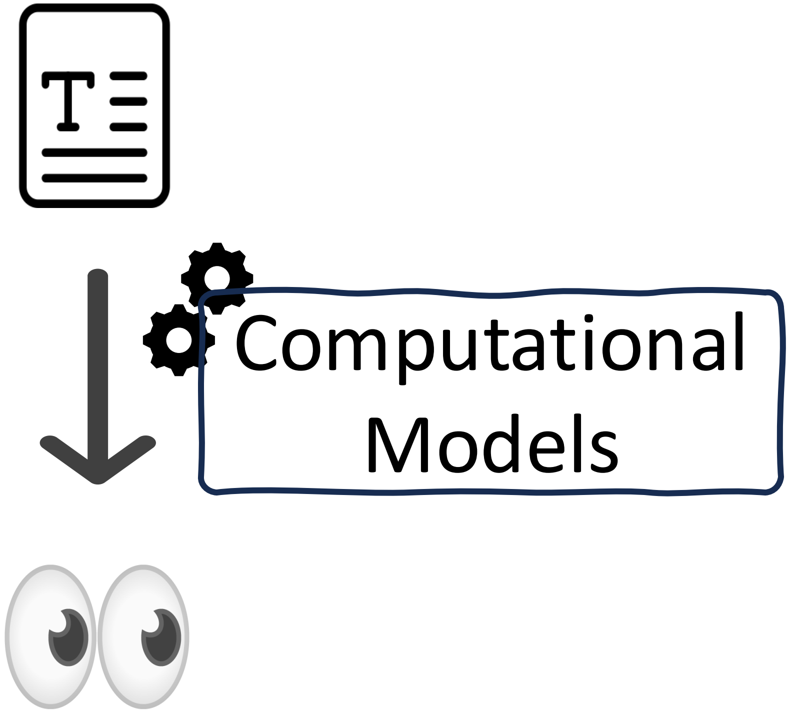
github.com/EyeBench

Shubi, Reich et al. (in prep)

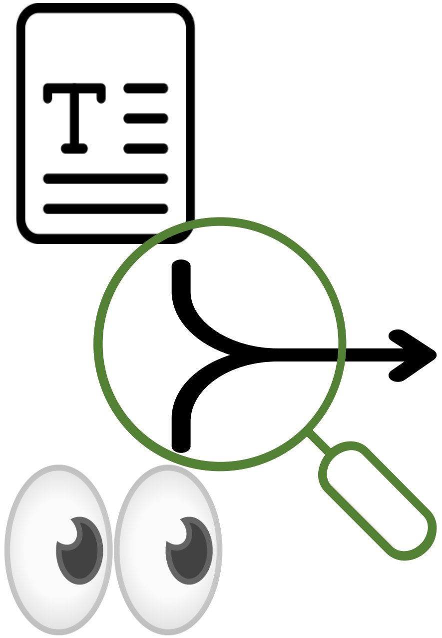
Next Up – EyeGenBench!

EyeGenBench: An Evaluation Framework for Models of Eye Movements in Reading

Coming Soon(ish)!



The Future: Interpretability



LLM ⇆ Human Attention Analysis

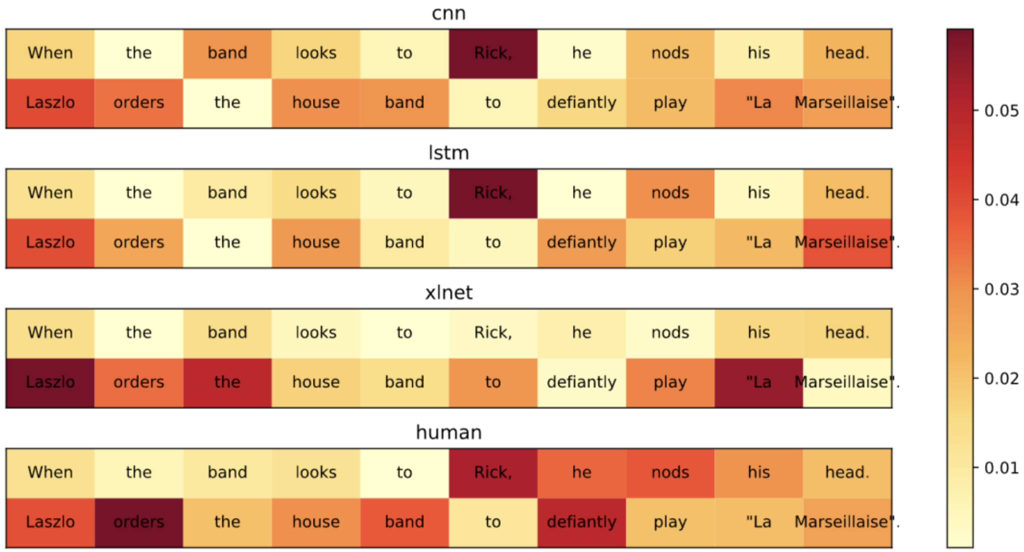
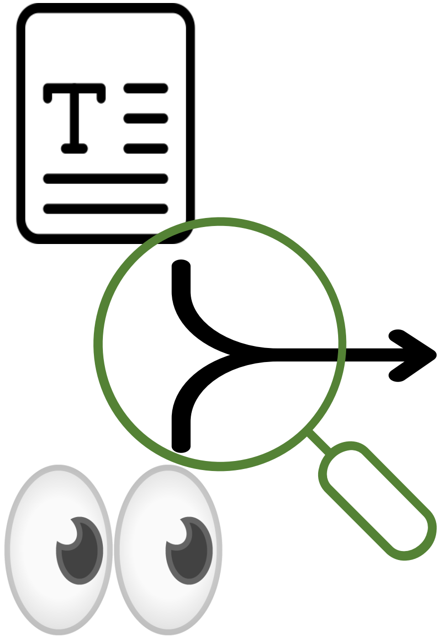


Figure 1: Example attention distributions of neural models (cnn, lstm, xlnet) and humans. Sood et al. (2020)

Sood et al. (2020), Eberle et al. (2022), Wang et al. (2024)

The Future: Interpretability



<u>**Explain and steer models**</u>

Identify reading strategies?

Do the models encode reader groups?

What properties are the models sensitive to?

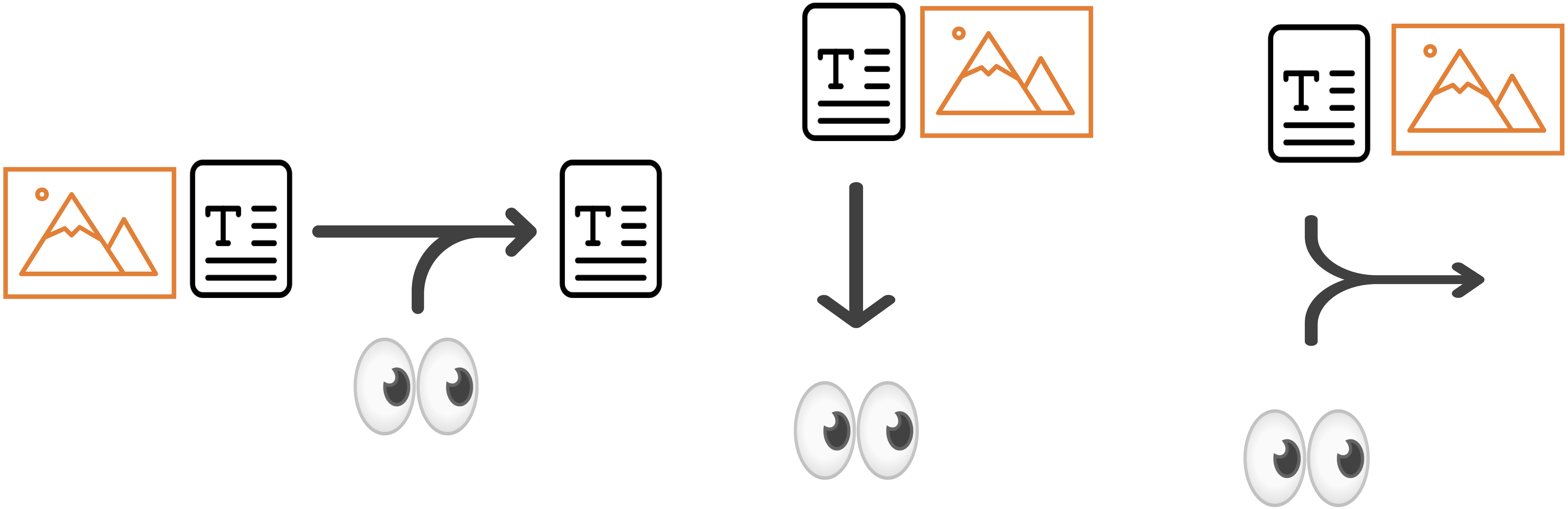
Simulate diverse readers and experiments?

The Future: Cognitive Alignment

Using **eye movements** to make **language models more human-like**

- Increasing the relevance of LMs as models of human linguistic processing
- Alignment with different target groups
- Alignment with other cognitive signals (e.g. the brain)
- Practical gains for NLP (e.g. resource efficiency)

The Future: Eye Tracking, Language and Vision



The Future: Eye Tracking, Language and Vision

Building on foundational work on **visual saliency**

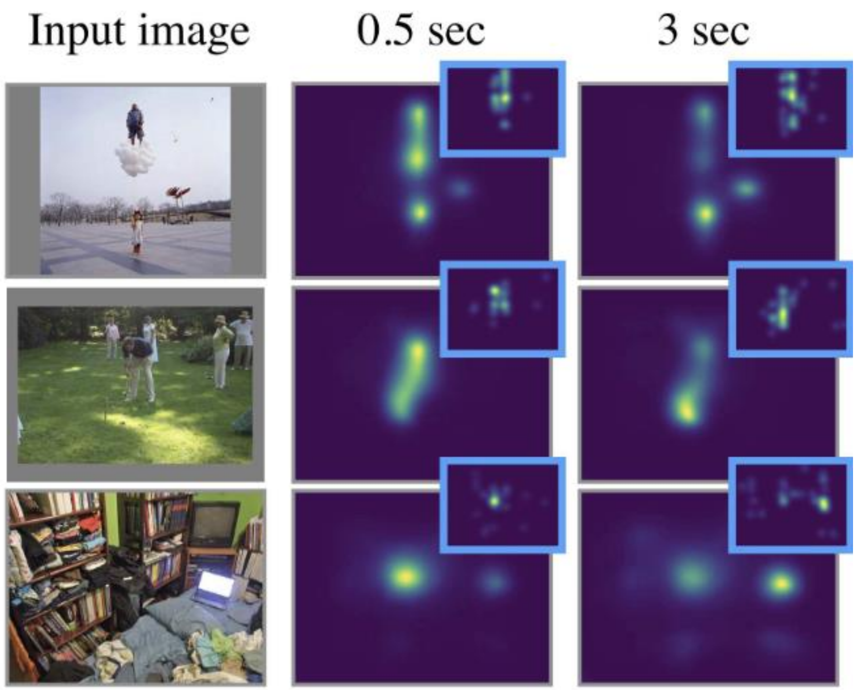
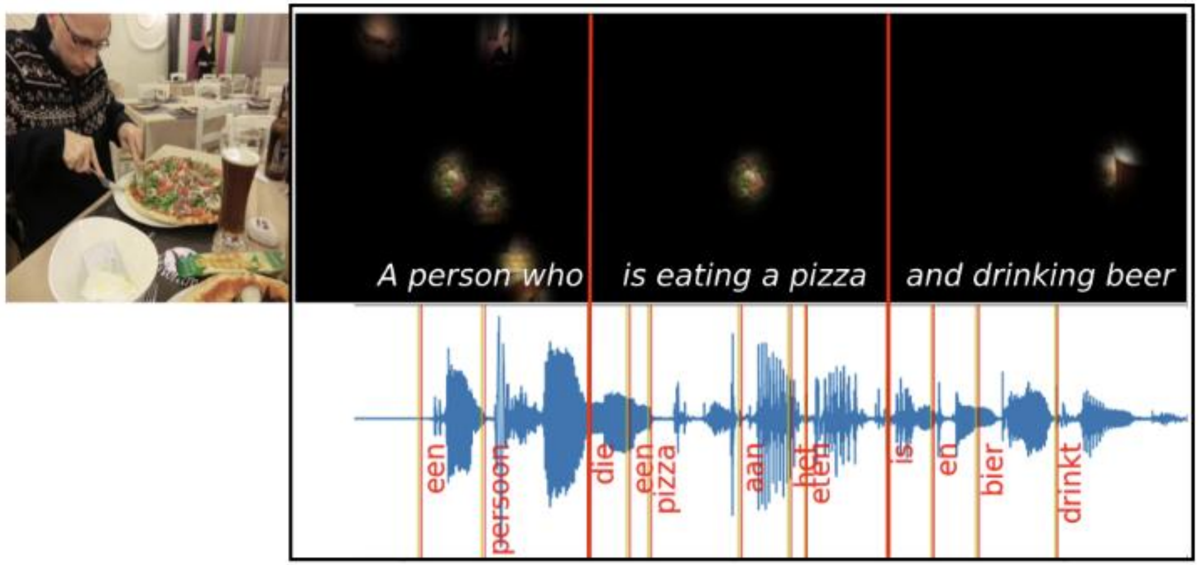


Image from Fosco et al. (2020)

The Future: Eye Tracking, Language and Vision

Image captioning

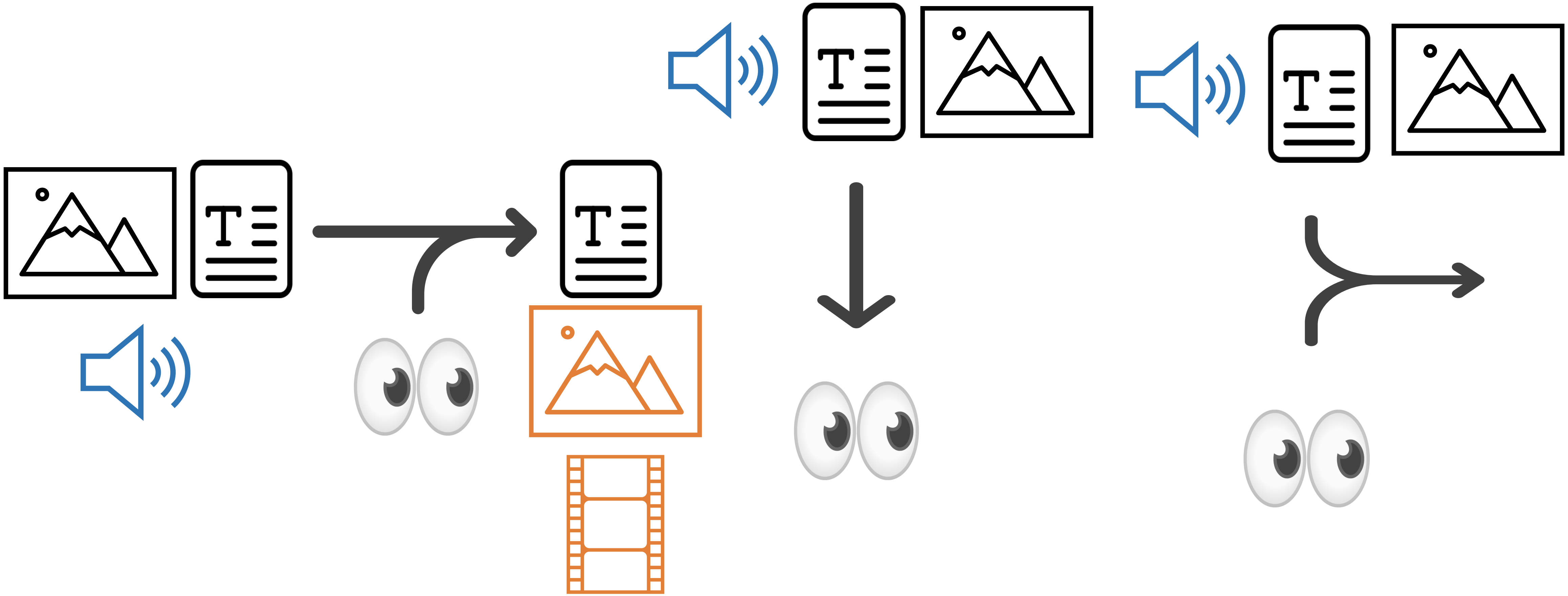
[Takmaz et al. (2020)](#)



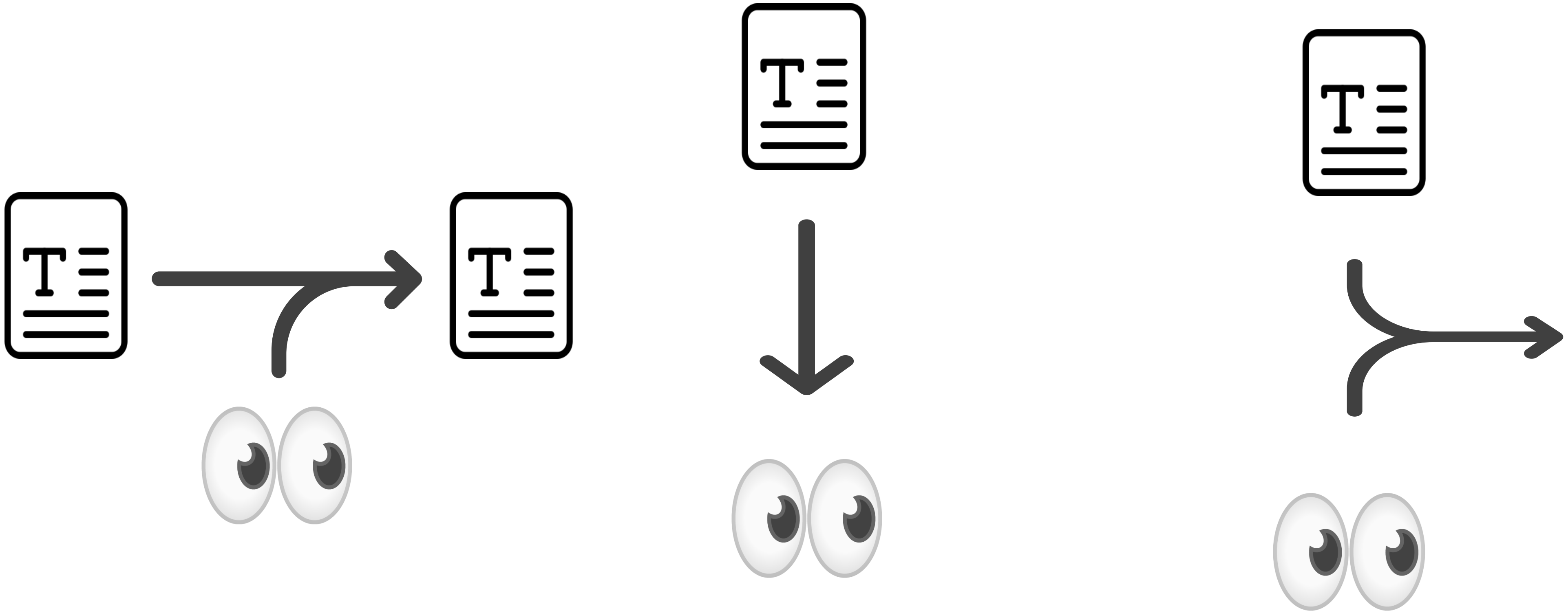
Visual question answering

[Sood et al. (2023)](#)

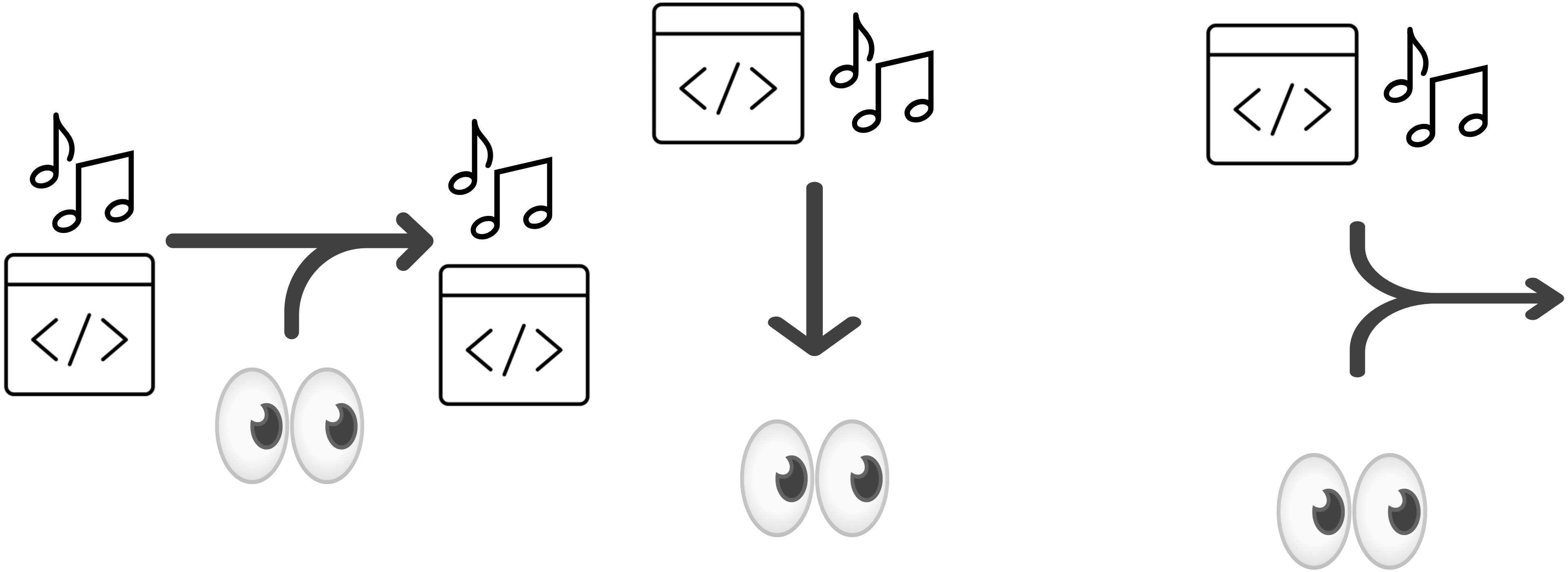
The Future: Eye Tracking, Language and Vision



The Future: Eye Tracking, Language and Vision



The Future: Eye Tracking, Language and Vision



The Future: Bias, Robustness and Fairness

Evaluating and ensuring model performance across groups

Reading Strategies

Educational Background

Age & Gender

Native Language & Language Skills

Human Centered NLP with Eye Movements Open Frontiers

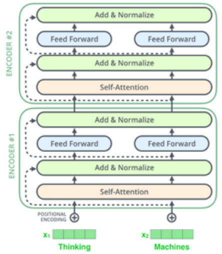
Information accessibility

Education and digital learning

Human-machine communication

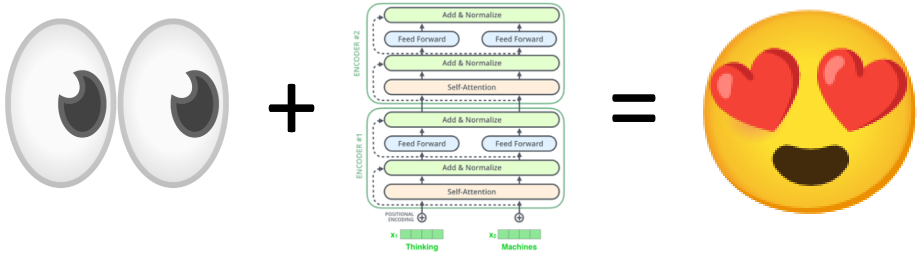
Interactive online systems

Integrate eye tracking data into YOUR research!

👀 + 

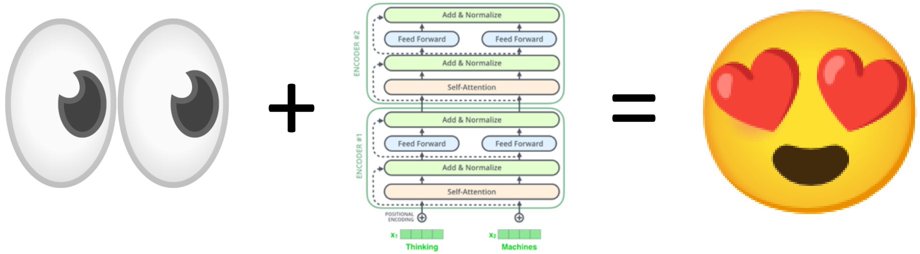
= 😍

What Now?



Slides and links to additional resources will be available on the [tutorial website]

What Now?



Join a new Discord channel on Eye Tracking and NLP

We thank Cui Ding, Jakub Dotlačil, Ece Takmaz and Shachar Frenkel for their support in preparing this tutorial!

TECHNION
Israel Institute of Technology

LaCC Lab

Department of Computational Linguistics

Universität Zürich[UZH]

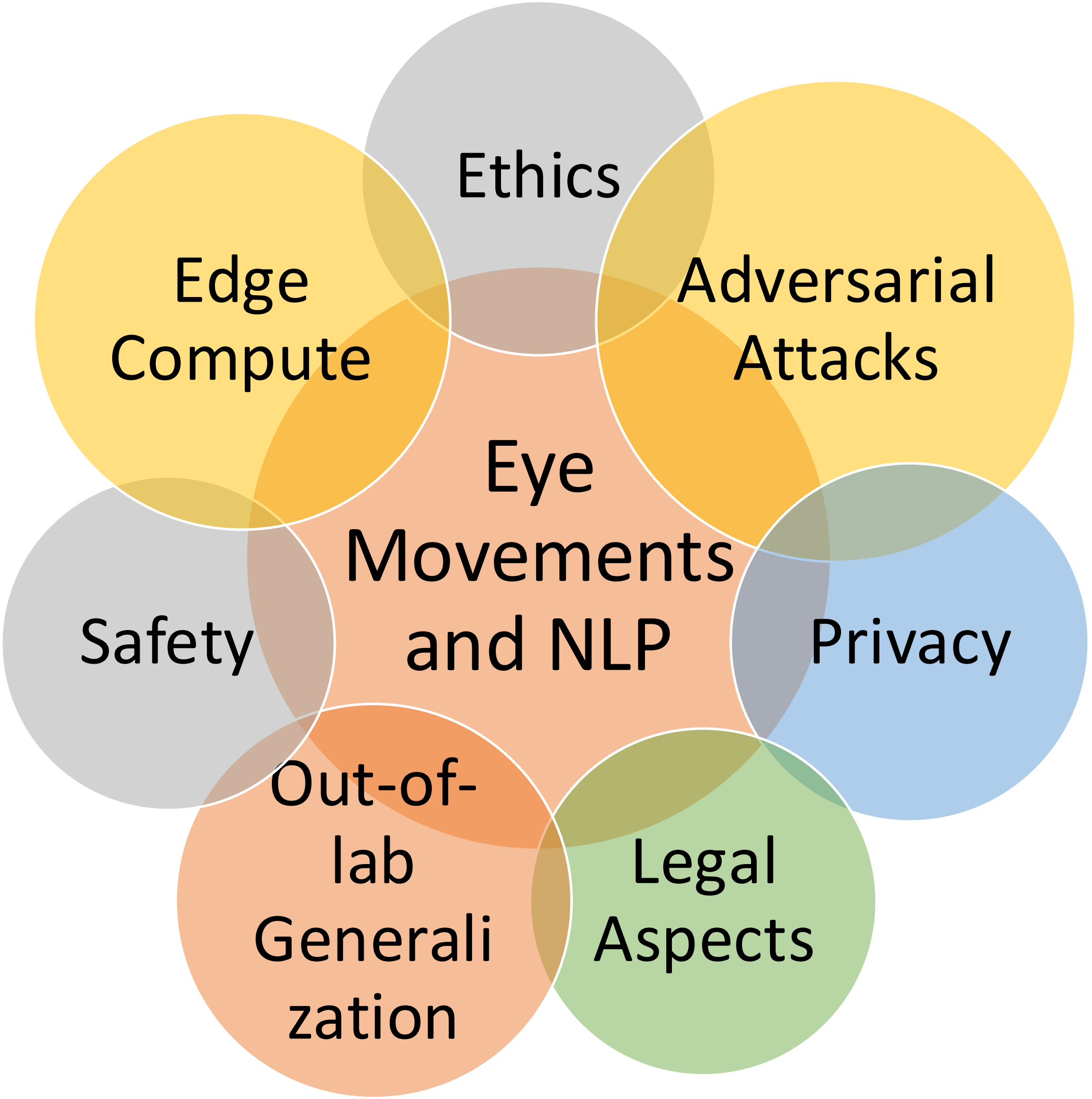
Eye Tracking and NLP Papers at ACL 2025

Session Details	Title
Monday, July 28 – 14:00-15:30 Room 1.61 – Session 3: IP-Orals	Déjà Vu? Decoding Repeated Reading from Eye Movements, Meiri et al.
Monday, July 28 – 18:00-19:30 Hall 4/5 – Session 5: IP-Posters	Beyond the Average Reader: the Reader Embedding Approach, Scozzaro et al.
Tuesday, July 29 – 10:30-12:00 Hall 4/5 – Session 7: IP-Posters	CogSteer: Cognition-Inspired Selective Layer Intervention for Efficiently Steering Large Language Models, Wang et al.
	Exploring the Effect of Nominal Compound Structure in Scientific Texts on Reading Times of Experts and Novices, Landwehr et al.
Wednesday, July 30 – 11:00-12:30 Hall 4/5 – Session 12: IP-Posters	Automatic detection of dyslexia based on eye movements during reading in Russian, Laurinavichyute et al.
	Decoding Reading Goals from Eye Movements, Shubi and Hadar et al.
	From Human Reading to NLM Understanding: Evaluating the Role of Eye-Tracking Data in Encoder-Based Models, Dini et al.
	ScanEZ: Integrating Cognitive Models with Self-Supervised Learning for Spatiotemporal Scanpath Prediction, Sood et al.
	Fine-Grained Spatio-Temporal Modeling of Reading Behavior, Re et al.

Ethical Considerations

- **Informed Consent** – Collect data only with IRB approval and written participant consent.
- **Privacy Protection** – Keep data anonymized; avoid storing information enabling user identification.
- **Responsible Use** – Only with explicit consent.
- **Bias Awareness** – Validate models for fairness, especially for L2 learners, cognitive/visual impairments.
- **Transparency** – Clearly communicate risks, limitations, and intended uses of predictive systems.

The Future 


Ethics
Edge Compute
Adversarial Attacks
Eye Movements and NLP
Safety
Privacy
Out-of-lab Generalization
Legal Aspects