

# Collapsing Distance: The Curse of Ground Truth in Computational Narrative Understanding

*Let's slow down and think about human label variation (HLV)*

Junbo Huang



*The pond*  
*Computational narrative understanding*



*The pond*  
*Computational narrative understanding*



Abundant compute

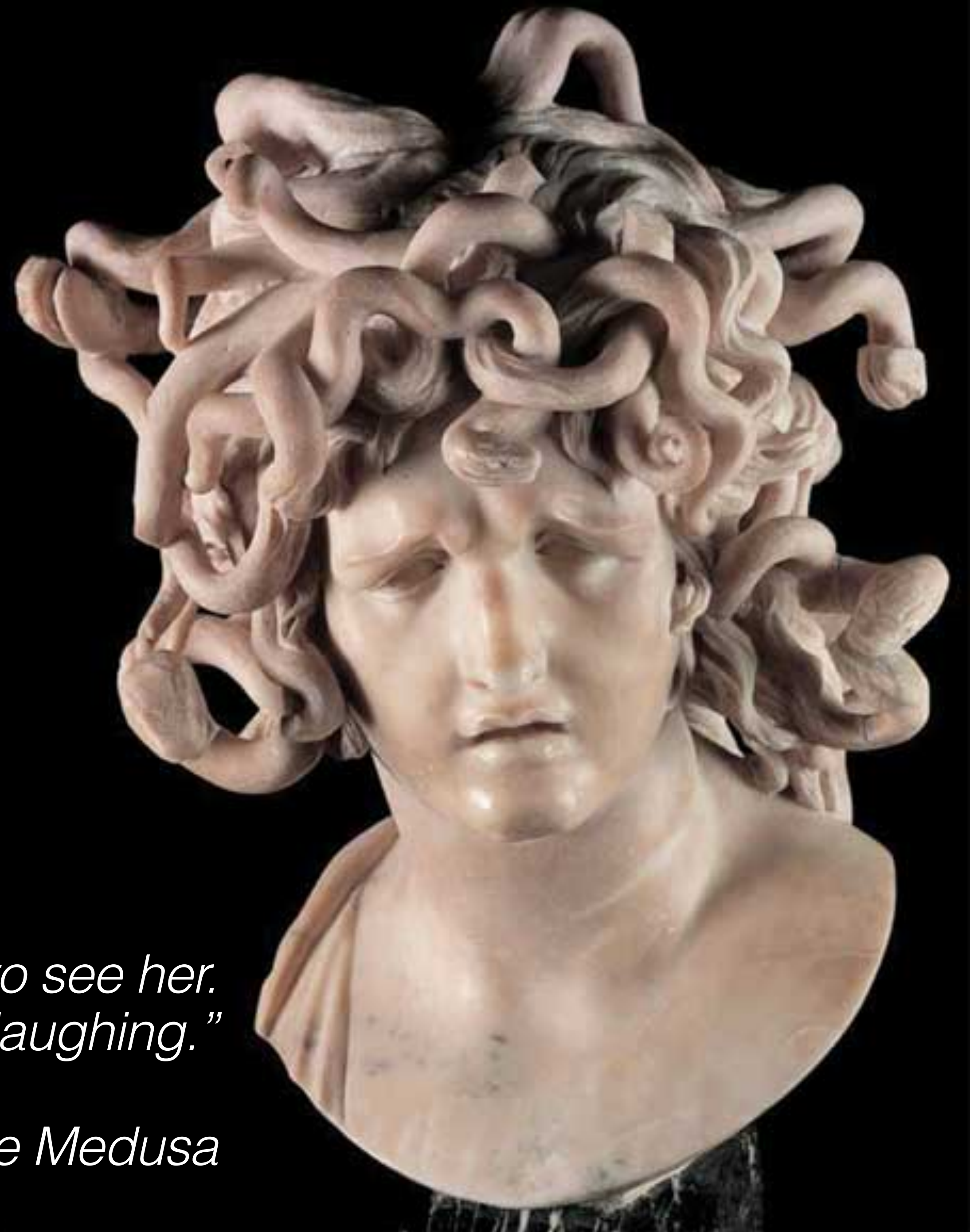
*Computational narrative understanding*

*The pond*

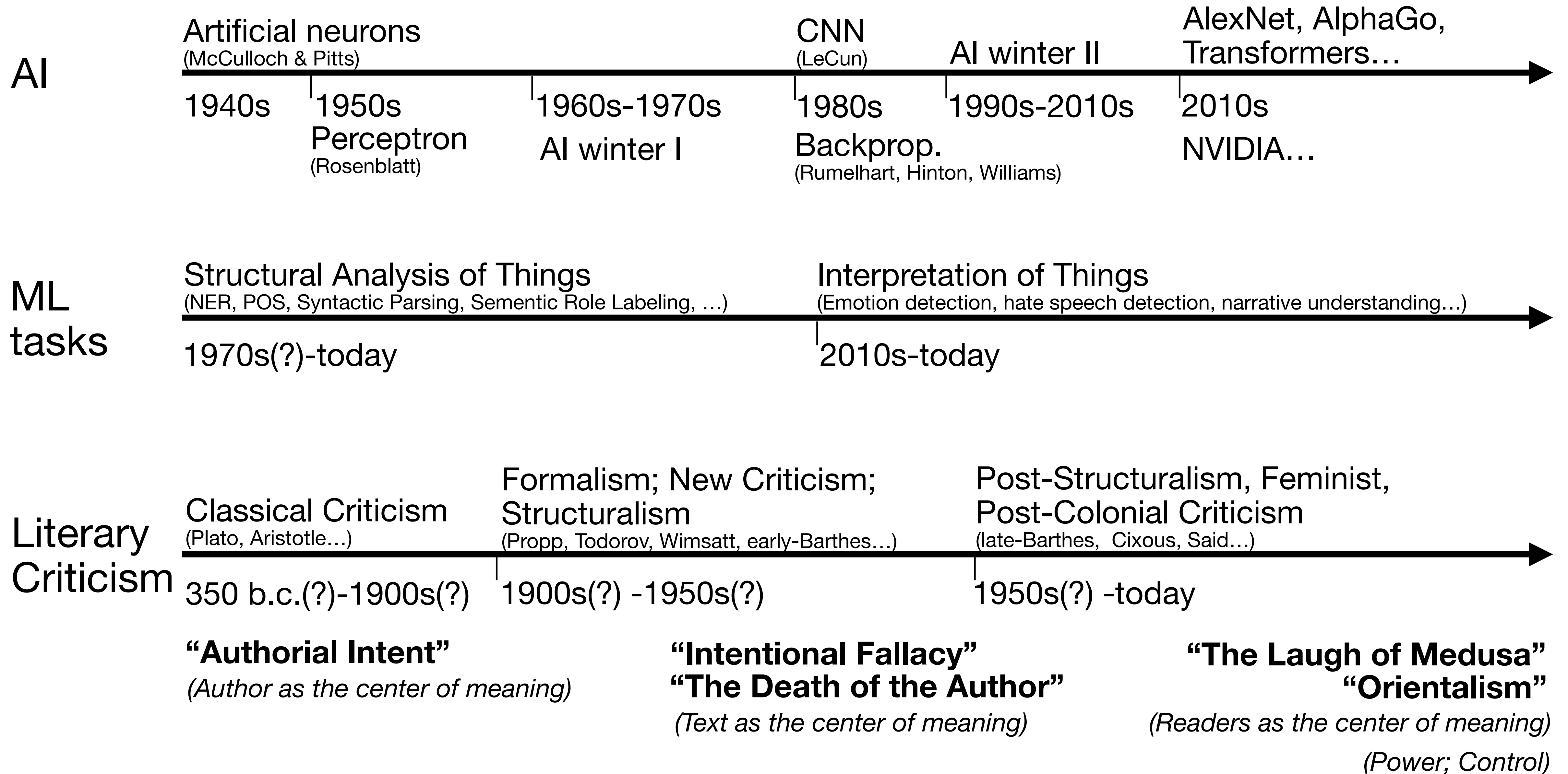
*Do meanings of narratives reside within the text itself?*

*“You only have to look at the Medusa straight on to see her.  
And she’s not deadly. She’s beautiful and she’s laughing.”*

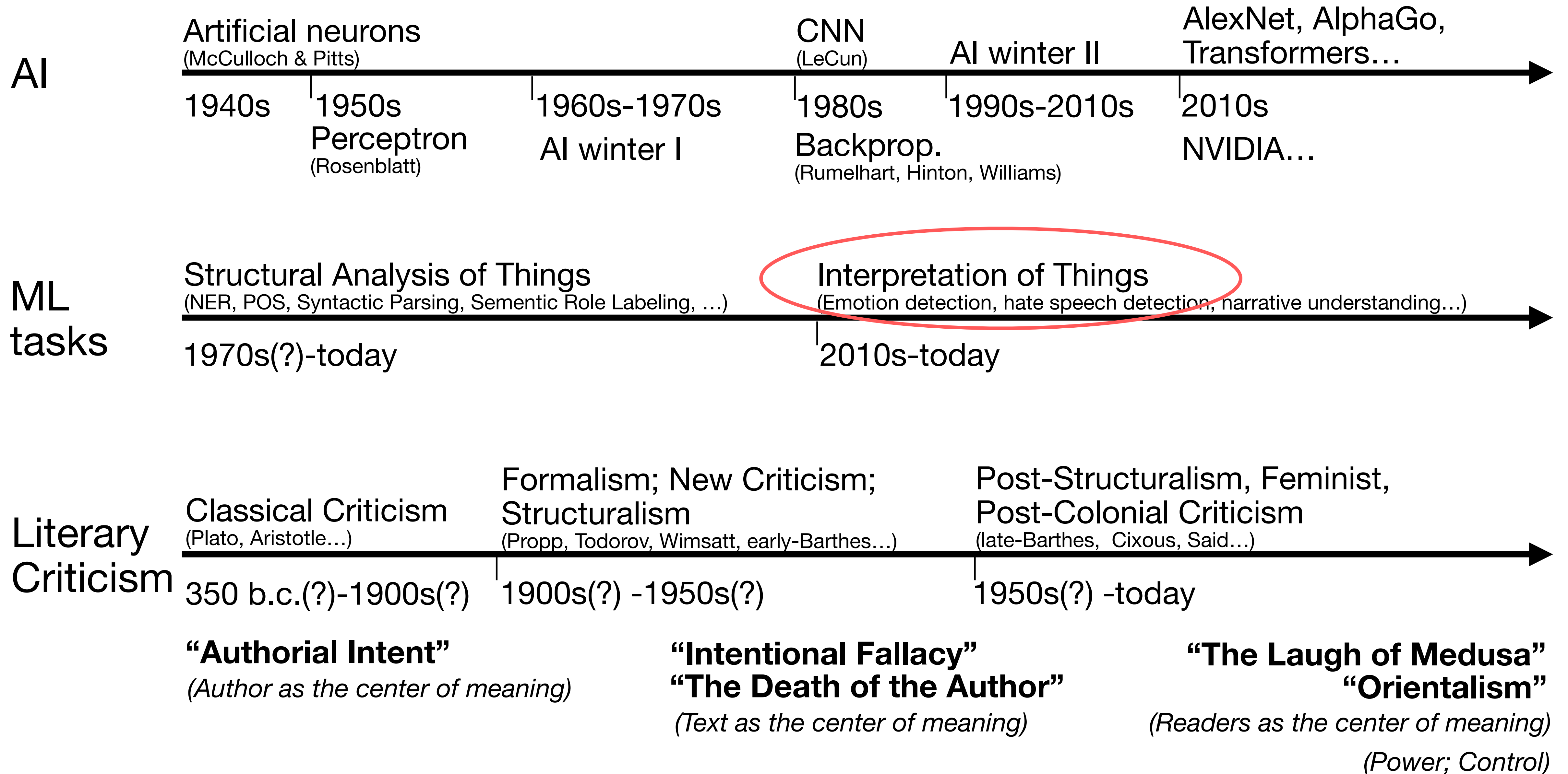
*— Hélène Cixous, The Laugh of the Medusa*



# Some *Rough* Timelines



# Some *Rough* Timelines



*“Michael Phelps' extraordinary career was formed by an incredible work ethic”  
—olympics.com*



Text source: <https://www.olympics.com/en/news/michael-phelps-training-regimen-workout-diet>

Image source: [https://i.guim.co.uk/img/media/29885232c95f4ce8bcc588bf800fbaa1bc650ed8/0\\_1092\\_1647\\_988/master/1647.jpg?width=1200&height=900&quality=85&auto=format&fit=crop&s=2f0861cda9c36236f6e4d0d2ab18d339](https://i.guim.co.uk/img/media/29885232c95f4ce8bcc588bf800fbaa1bc650ed8/0_1092_1647_988/master/1647.jpg?width=1200&height=900&quality=85&auto=format&fit=crop&s=2f0861cda9c36236f6e4d0d2ab18d339)

*“Usain Bolt is a genetic freak because being 6’5” tall means he shouldn’t be able to accelerate at the speed he does.”*  
—BBC



# Narrative Production

What do computational models of narratives extract?

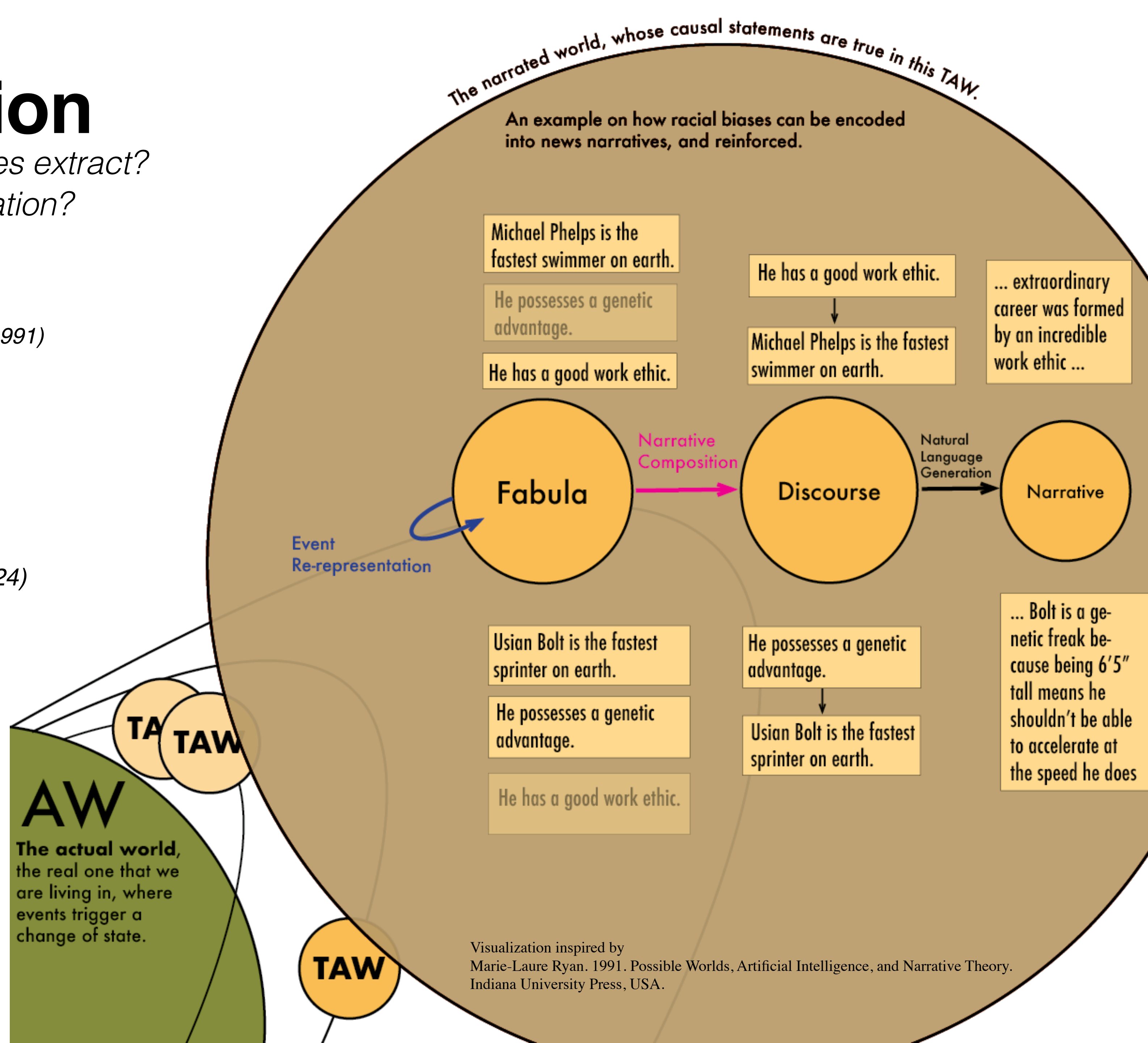
Who makes sense of the extracted information?

## Possible Worlds in Narratology (Ryan, 1991)

1. Actual worlds (AW)
2. Textual actual worlds (TAW)
3.  $d(AW, TAW) \approx \text{bias ?}$

## News Narration as functions (Huang, 2024)

1. Event re-representation
2. Narrative composition (Gervás, 2013)
3. Natural language generation



- Pablo Gervás. 2013. Narrative Composition: Achieving the Perceived Linearity of Narrative. In Proceedings of the 14th European Workshop on Natural Language Generation, pages 103–104, Sofia, Bulgaria. Association for Computational Linguistics.
- Junbo Huang and Ricardo Usbeck. 2024. Narration as Functions: from Events to Narratives. In Proceedings of the 6th Workshop on Narrative Understanding, pages 1–7, Miami, Florida, USA. Association for Computational Linguistics.

# Narrative Production

What do computational models of narratives extract?

Who makes sense of the extracted information?

Where do people look at?

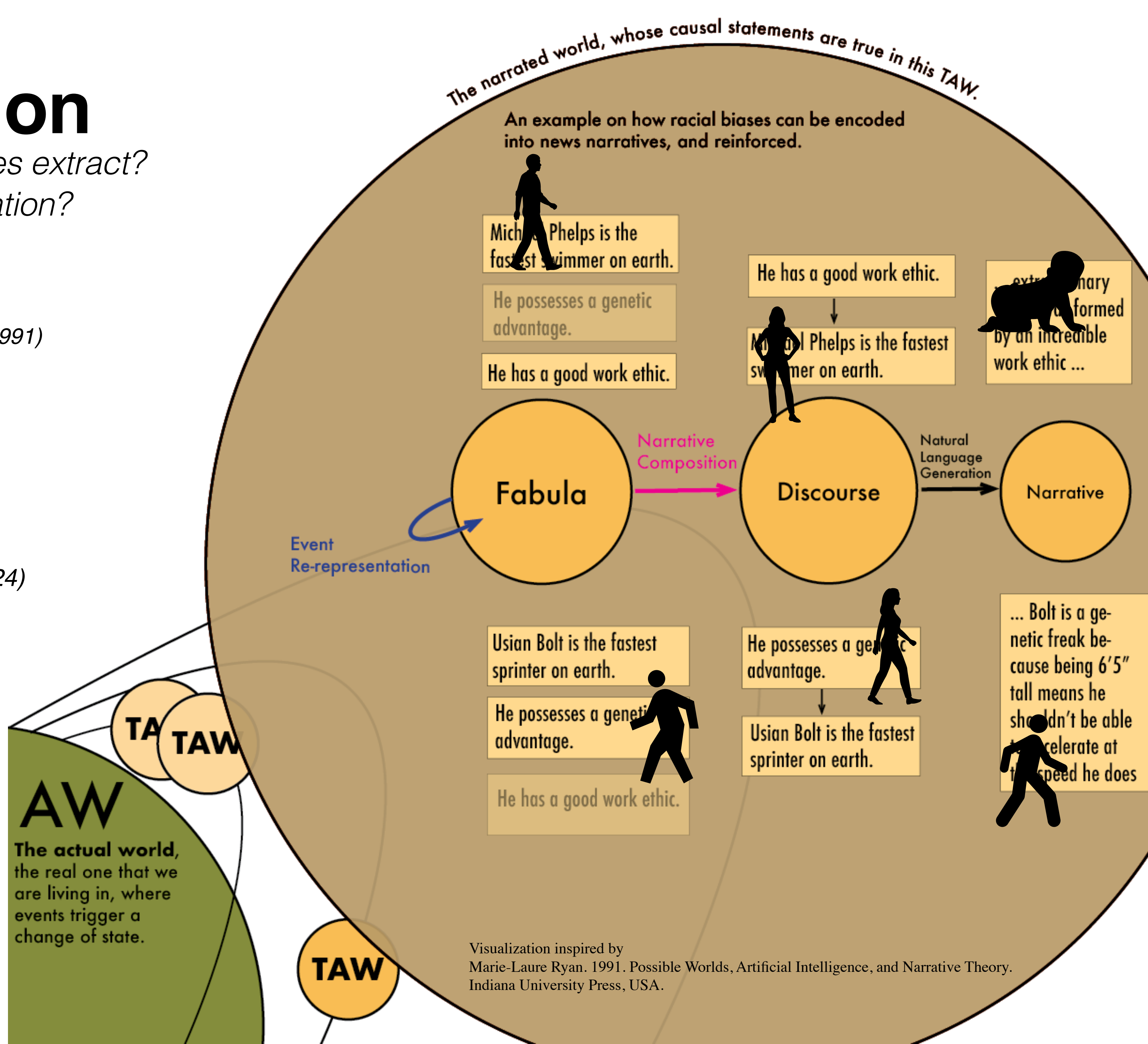
## Possible Worlds in Narratology (Ryan, 1991)

1. Actual worlds (AW)
2. Textual actual worlds (TAW)
3.  $d(AW, TAW) \approx \text{bias}$  ?

## News Narration as functions (Huang, 2024)

1. Event re-representation
2. Narrative composition (Gervás, 2013)
3. Natural language generation

## Distribution of interpretation



Visualization inspired by Marie-Laure Ryan. 1991. Possible Worlds, Artificial Intelligence, and Narrative Theory. Indiana University Press, USA.

- Pablo Gervás. 2013. Narrative Composition: Achieving the Perceived Linearity of Narrative. In Proceedings of the 14th European Workshop on Natural Language Generation, pages 103–104, Sofia, Bulgaria. Association for Computational Linguistics.
- Junbo Huang and Ricardo Usbeck. 2024. Narration as Functions: from Events to Narratives. In Proceedings of the 6th Workshop on Narrative Understanding, pages 1–7, Miami, Florida, USA. Association for Computational Linguistics.

# Narrative Definitions

*Formal structure = being objective?*

## Narratives as formal structure

- *Shklovsky (1923)*: fabula vs syuzhet
- *Wimsatt (1946)*: structure is within the text
- *Barthes (1966)*: nuclei, catalysers
- *Propp (1968)*: character roles
- *Todorov (1969)*: equilibrium, disruption, equilibrium
- *Genette (1980)*: voice (narrative levels), mood (point of views), tense (event relations)

## Narratives as phenomena

- *Foucault (1979), Barthes (1986)*: power & control in narrative production
- *Cixous (1976)*: Écriture féminine (breaking from the traditional masculine styles of writing)
- *Said (1978)*: the west's imperialistic and colonial portrait of the east
- *Ryan (2014)*: storyworlds as readers' mental models when engaging with narratives

# Narrative Definitions

*Formal structure = being objective?*

## Narratives as formal structure

*meanings lay within the text itself*

- *Shklovsky (1923)*: fabula vs syuzhet
- *Wimsatt (1946)*: structure is within the text
- *Barthes (1966)*: nuclei, catalysers
- *Propp (1968)*: character roles
- *Todorov (1969)*: equilibrium, disruption, equilibrium
- *Genette (1980)*: voice (narrative levels), mood (point of views), tense (event relations)

## Narratives as phenomena

*meanings lay beyond the text itself*

- *Foucault (1979), Barthes (1986)*: power & control in narrative production
- *Cixous (1976)*: Écriture féminine (breaking from the traditional masculine styles of writing)
- *Said (1978)*: the west's imperialistic and colonial portrait of the east
- *Ryan (2014)*: storyworlds as readers' mental models when engaging with narratives

# Narrative Interpretation

*Why do annotators disagree? Can they disagree?*

*Do you want to keep these interpretive variations in your dataset?*

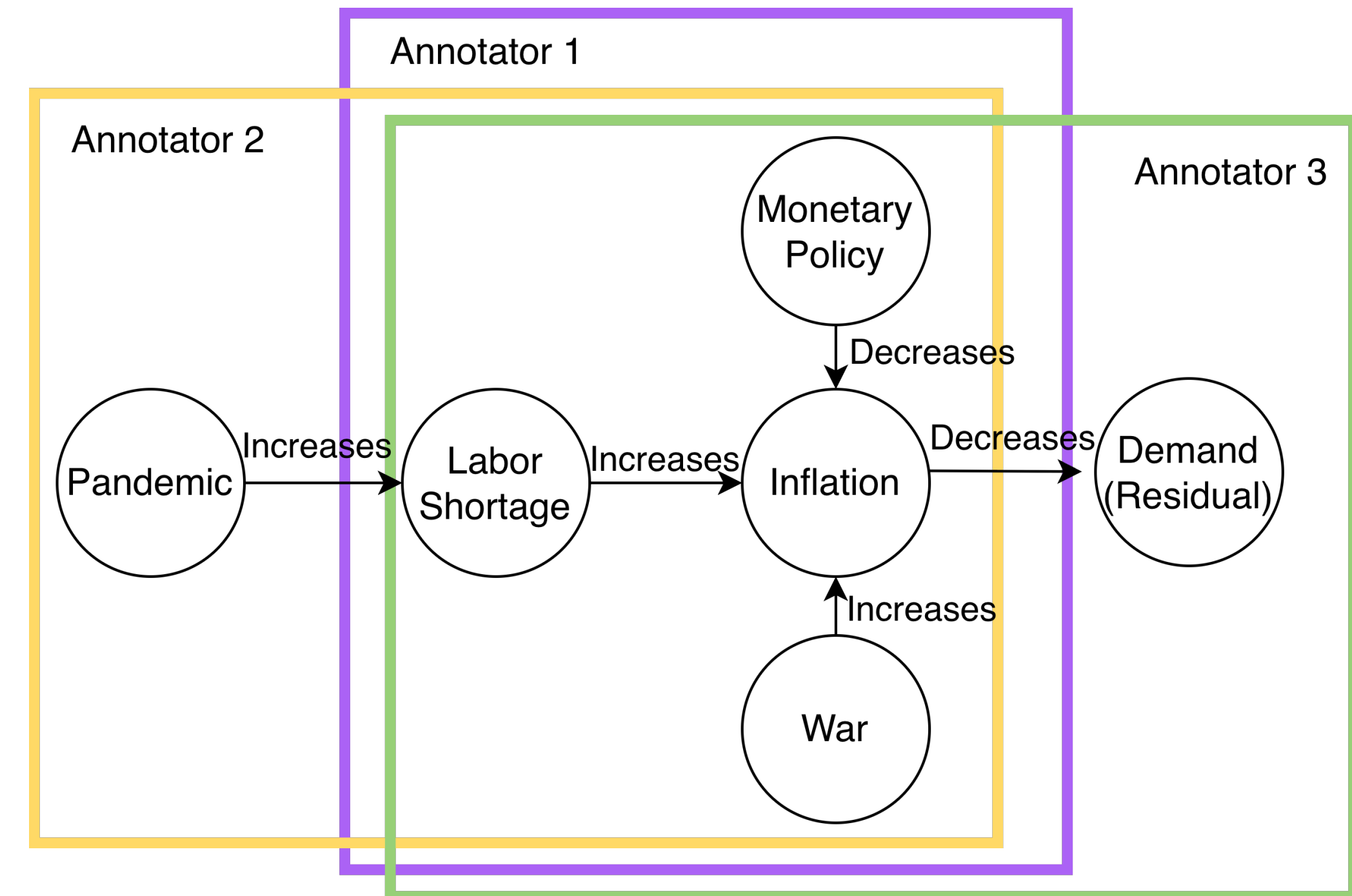
*Do you want your model to produce these variations?*

## Narrative annotation: human interpreting texts

1. There are variations in narrative annotation
  - *“What is told to cause inflation?”* (Huang, 2026)
  - *“Does this text contain a story?”* (Antoniak, 2024)
2. Rethinking variation in your narrative annotation tasks
  - *Is variation noise? or signal?*

## Narrative extraction: algorithms “interpreting” text

1. Extracting events, characters, times, places, narrative levels, point of views
2. Symbolic approach v.s. neural approach v.s. hybrid



• Junbo Huang, Max Weing, Ulrich Fritsche, Ricardo Usbeck. 2026. [From variance to invariance: Qualitative content analysis for narrative graph annotation](#), In *Proceedings of the Language Resources and Evaluation Conference*, Palma de Mallorca, Spain. Association for Computational Linguistics.

• Maria Antoniak, Joel Mire, Maarten Sap, Elliott Ash, and Andrew Piper. 2024. [Where Do People Tell Stories Online? Story Detection Across Online Communities](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7104–7130, Bangkok, Thailand. Association for Computational Linguistics.

# Human Label Variation (Plank, 2022)

*What has been long considered “noise”*

Types of annotation variations:

1. *Annotation error*
2. *Subjectivity* ✿
3. *Disagreement* ✿
4. *Multiple plausible answers* ✿



In ML, often they are viewed as:

*“Noise”*

✿ *genuine variation*

**This work shows** *the impact of removing different readings of the same text.*

- *What does inter-annotator agreement measure?*
- *What does distance metric loss optimize toward?*
- *What do evaluation metrics measure?*

# Collapsed Reliability

*Majority voting favors a single, fixed ground truth reading of narratives*

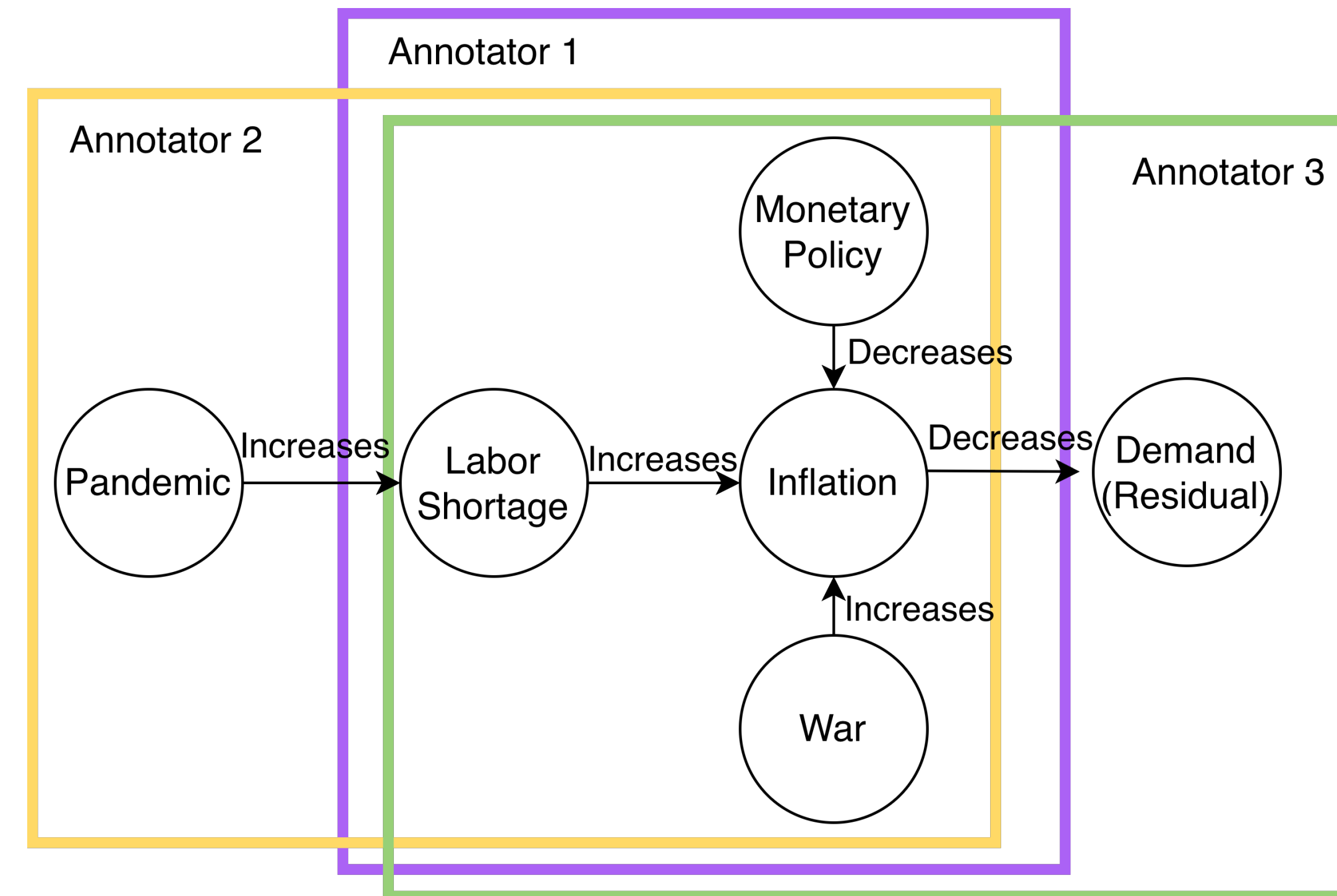
## Inter-annotator agreement (IAA)

- *Cohen's kappa*
- *Fleiss's kappa*
- *Krippendorff's alpha*

*(weighted sum of pairwise distances between annotations.)*

## Genuine variations still exist in reliable annotations

- *Provide unaggregated labels*
- *Collect meta-data (e.g., sociodemographics)*
- *Reduce annotation errors*
  - *Rigorous annotation methodology (e.g., qualitative content analysis)*

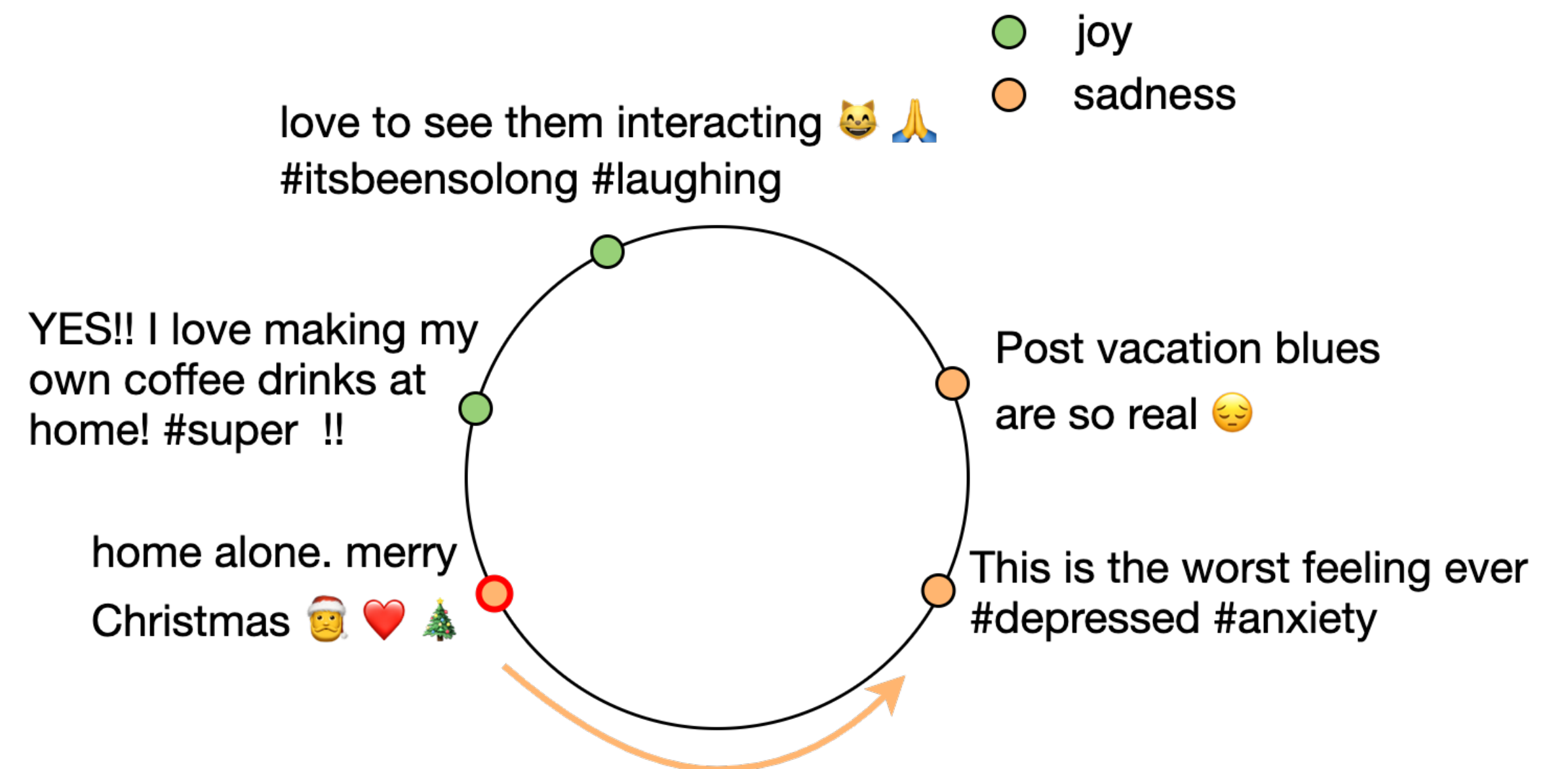
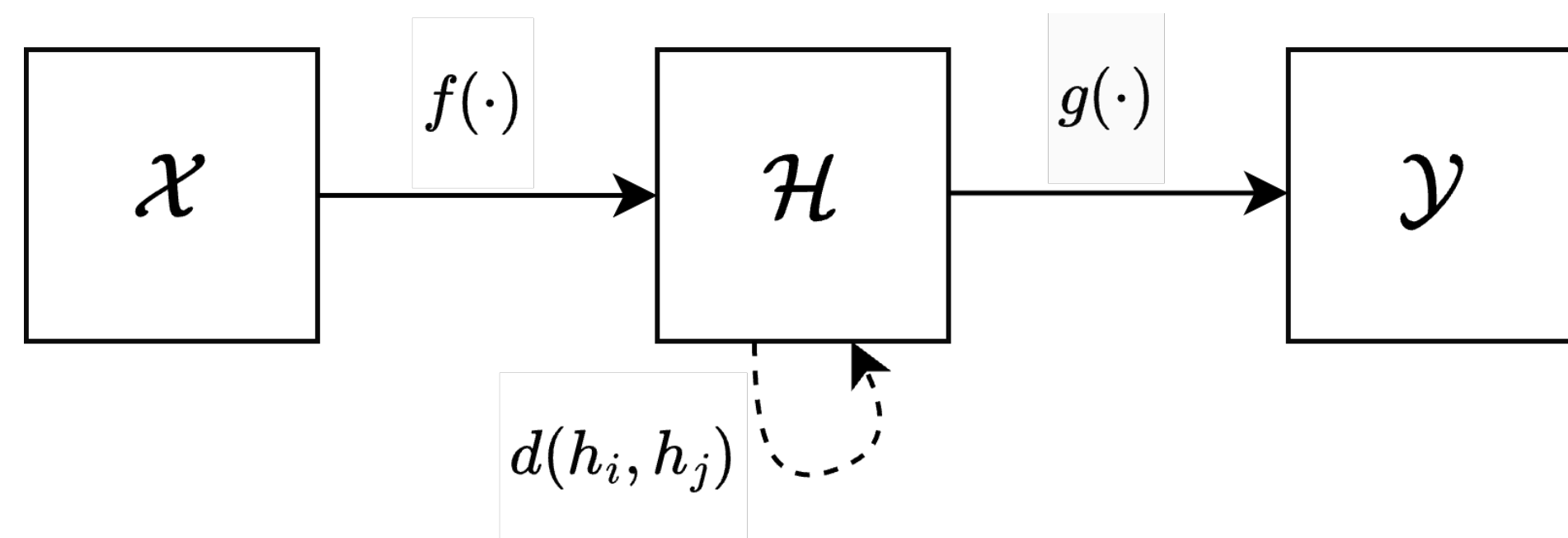


# Collapsed Embedding Space

Can representation learning algorithms capture abstract and complex phenomena with majority voting?

## Distance metric learning with Siamese Neural Networks

- Goal: to learn a latent space  $\mathcal{H}$  such that similar pairs (positives) are pulled together while dissimilar pairs (negatives) are pushed apart.
- Learn to structure  $\mathcal{H}$  such that  $d(h_i, h_j)$  reflects  $d(y_i, y_j)$
- Example: FaceNet, Sentence-BERT, CLIP



# Collapsed Model Evaluation

*Evaluation metrics are biased towards the dominant perspective*

## Careful interpretation of derived scores (e.g., F1)

- *F1=0.99?*
- *Model performs very well (in reproduce the dominant voice)*

## Soft metrics (Rizzi, 2024)

- *Cross entropy*
- *KL divergence*
- *...*

# Final Remarks

*What has been long considered “noise”*

*Slow down,  
new methods needed for annotation evaluation,  
model training and model evaluation*

## Literary Studies

*Top-down theoretical hypotheses  
The teacher  
The helper*

## Machine learning

*Bottom-up computational methods  
The student  
The hero and the villain*

*I am sorry, but narratives are not processed products  
like cars...*