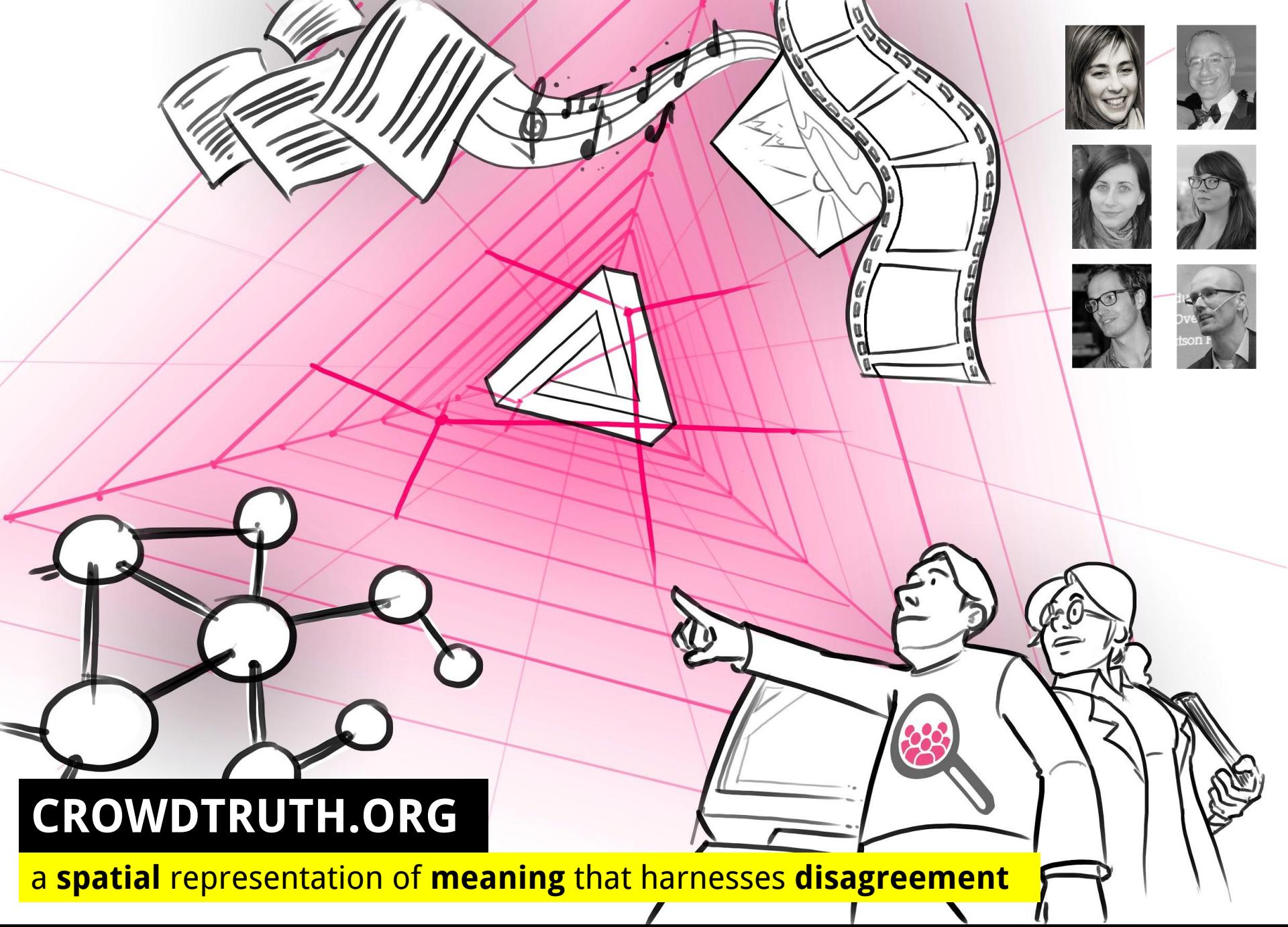
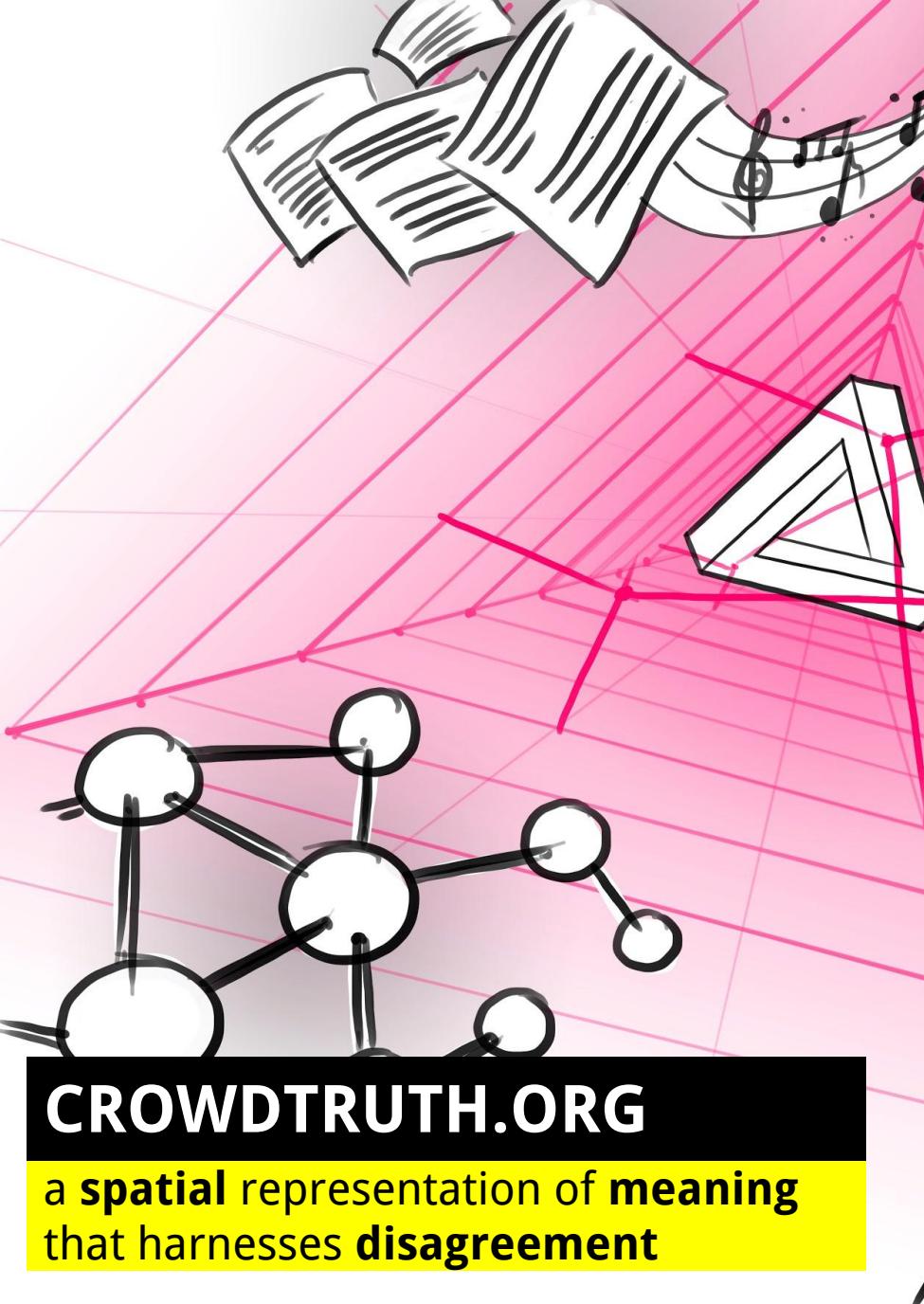


CROWDTRUTH TUTORIAL

Part I: Introduction to CrowdTruth

Lora Aroyo, Anca Dumitache, Oana Inel, Chris Welty





a human computation
(crowdsourcing) approach to:

- gather **diversity** of **perspectives** & **opinions** from **crowds** & **niches**
- expand **expert vocabularies** with such perspectives and opinions
- gather **new type of gold standard** for machines that considers diversity and ambiguity



COMFORT ZONE

7 MYTHS ABOUT HUMAN ANNOTATION

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty



ONE TRUTH

7 MYTHS ABOUT HUMAN ANNOTATION

One truth: knowledge acquisition assumes one correct interpretation for every example

“Truth is a Lie: 7 Myths about Human Annotation”, *AI Magazine* 2014, L. Aroyo, C. Welty



ALL EXAMPLES EQUAL

7 MYTHS ABOUT HUMAN ANNOTATION

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty



DISAGREEMENT BAD

7 MYTHS ABOUT HUMAN ANNOTATION

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty



EXPERTS RULE

7 MYTHS ABOUT HUMAN ANNOTATION

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

Experts rule: knowledge is captured from domain experts



ONE IS ENOUGH

7 MYTHS ABOUT HUMAN ANNOTATION

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

Experts rule: knowledge is captured from domain experts

One is enough: knowledge by a single expert is sufficient



BINARY WORLD

7 MYTHS ABOUT HUMAN ANNOTATION

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

Experts rule: knowledge is captured from domain experts

One is enough: knowledge by a single expert is sufficient

Detailed explanations help: if examples cause disagreement - add instructions

Once done, forever valid: knowledge is not updated; new data not aligned with old

DOES THIS SENTENCE EXPRESS TREATS RELATION?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated
with **CHLOROQUINE** for decades.

DOES THIS SENTENCE EXPRESS TREATS RELATION?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

DOES THIS SENTENCE EXPRESS TREATS RELATION?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

WHAT DO EXPERTS SAY?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.



For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.



Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



WHAT DOES A LAY ANNOTATOR SAY?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.



For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.



Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



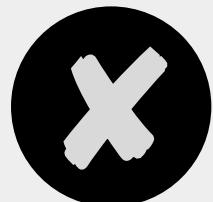
WHAT DOES ANOTHER LAY ANNOTATOR SAY?

Treats: Chloroquine, Malaria

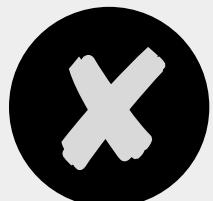
Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.



For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.



Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



WHAT DOES A THIRD LAY ANNOTATOR SAY?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.



For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.



Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



WHAT DOES THE CROWD SAY?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

95%

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

75%

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

50%

Intuition: This is better

WHAT DOES THE CROWD SAY?

Treats: Chloroquine, Malaria

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

95%

BETTER

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

75%

WORSE

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

50%

This one isn't utterly wrong



COMFORT ZONE

Disrupted

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

Experts rule: knowledge is captured from domain experts

One is enough: knowledge by a single expert is sufficient

Detailed explanations help: if examples cause disagreement - add instructions

Once done, forever valid: knowledge is not updated; new data not aligned with old

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty



DISRUPTING THE COMFORT ZONE

Encourage Disagreement in Annotation Tasks & Develop New Quality Metrics

“Truth is a Lie: 7 Myths about Human Annotation”, *AI Magazine* 2014, L. Aroyo, C. Welty

CROWD TASK



Medical Relation Extraction

powered by
CrowdTruth 
The framework for crowdsourcing ground truth data

1

In the following sentence:

Sentence: _____

Among 56 subjects reporting to a clinic with symptoms of **malaria**, 53 (95%) had ordinarily effective levels of **chloroquine** in blood.

2

Is **chloroquine** related to **malaria**? Choose all that apply.

Treats

Diagnosed By

Causes

Location

✓ Manifestation

Contraindicates

✓ Associated With

Is A

Part Of

✓ Symptom

✓ Other

None

SYMPOTM: Deviation from normal function indicating the presence of disease or abnormality, e.g. pain is a symptom of a broken arm.

WORKER VECTOR FOR A SENTENCE

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



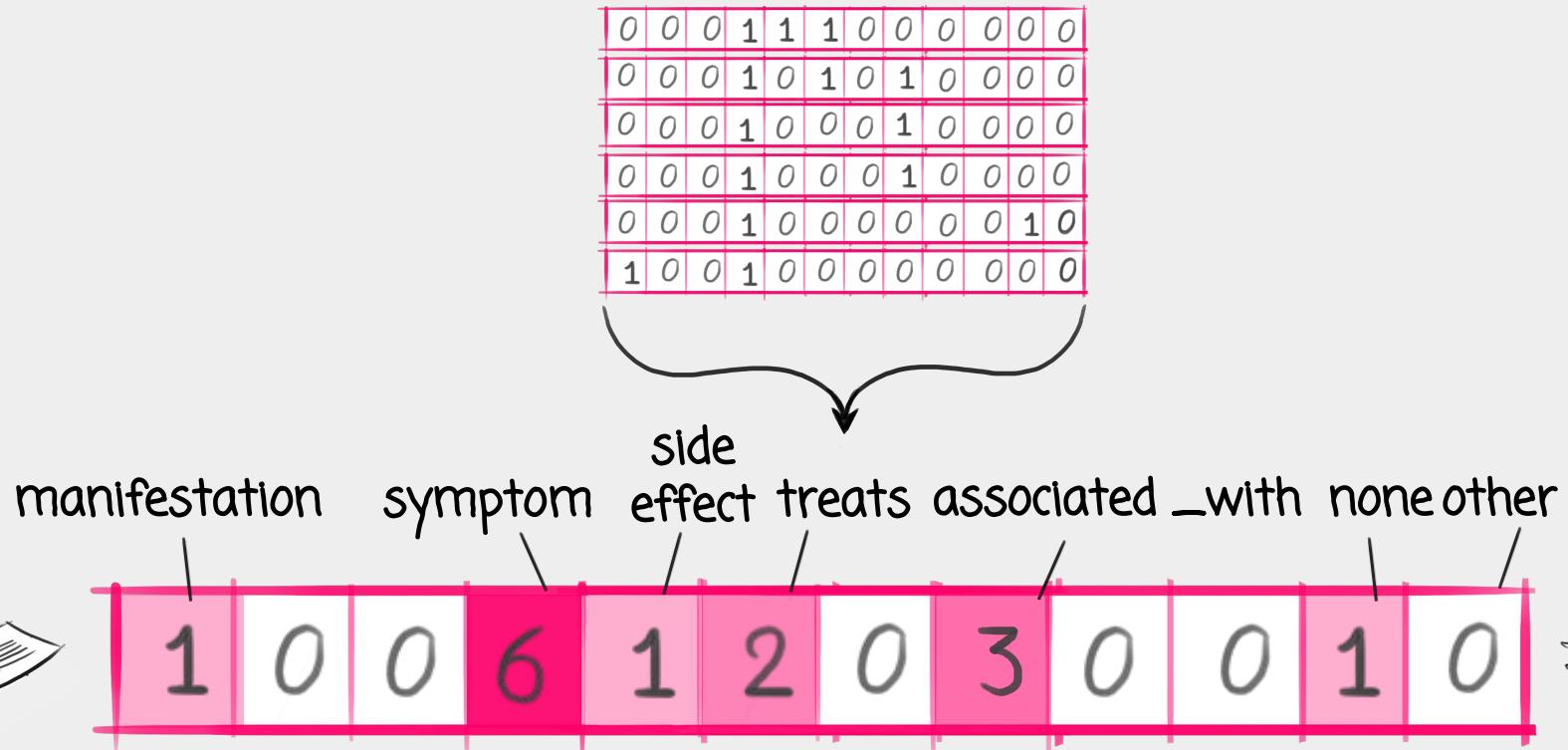
MANY WORKERS FOR THE SAME SENTENCE

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

	symptom		treats		associated _with	other					
	0	0	0	1	1	1	0	0	0	0	0
	0	0	0	1	0	1	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0	0
	0	0	0	1	0	0	0	0	0	0	1
	1	0	0	1	0	0	0	0	0	0	0

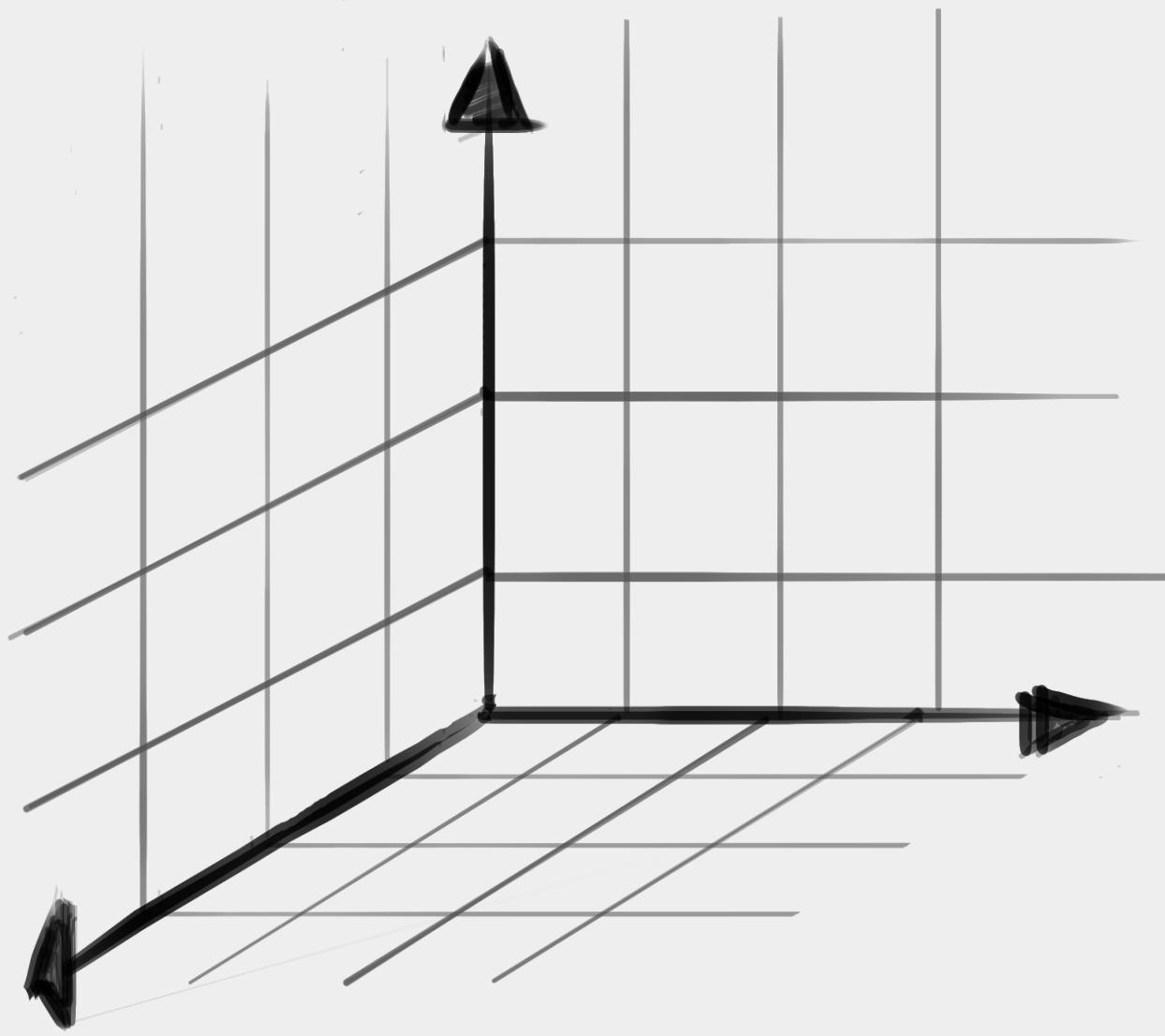
ALL WORKER VECTORS AGGREGATED IN A SENTENCE VECTOR

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



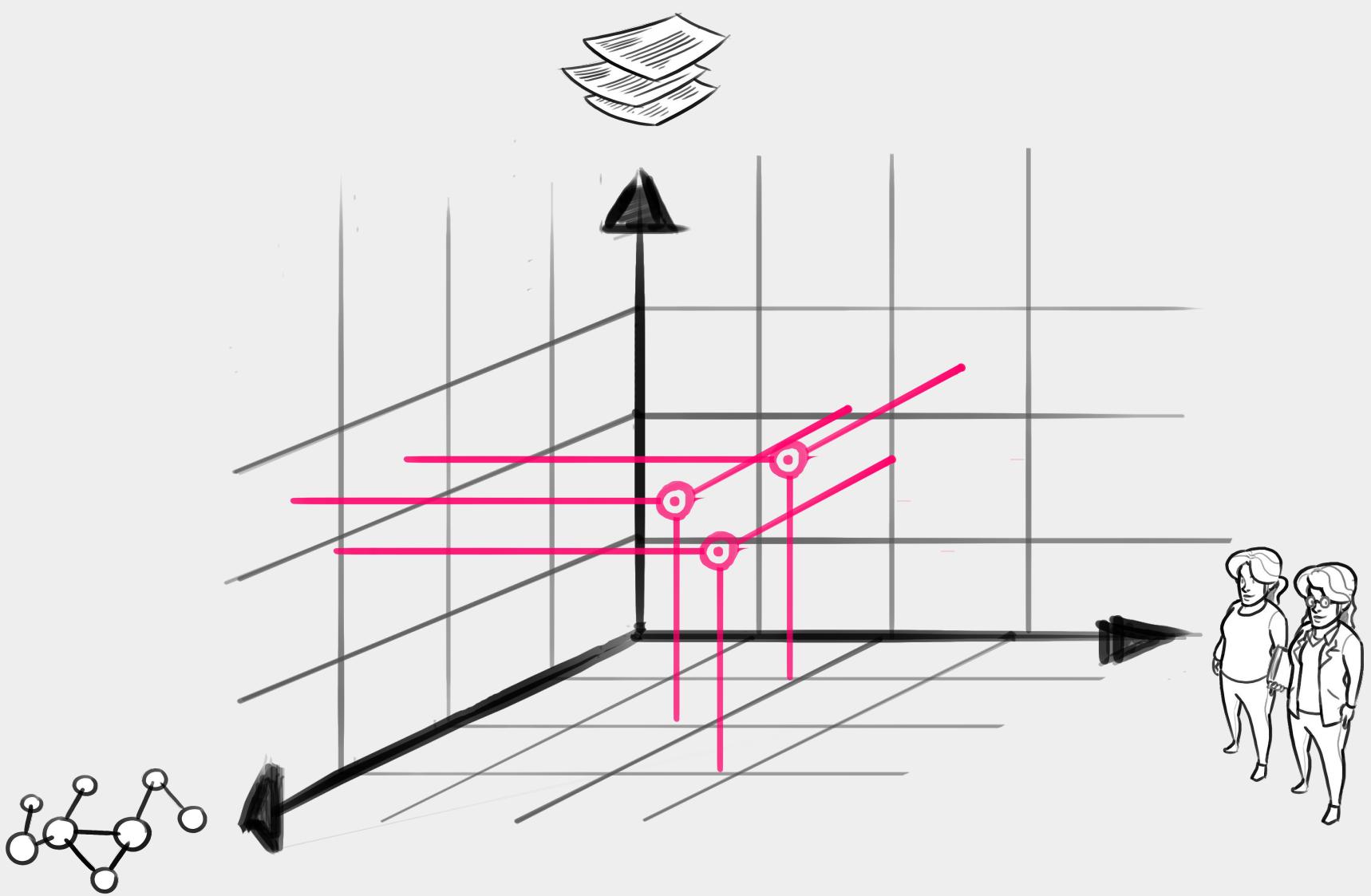
SENTENCE VECTORS FOR DIFFERENT SENTENCES





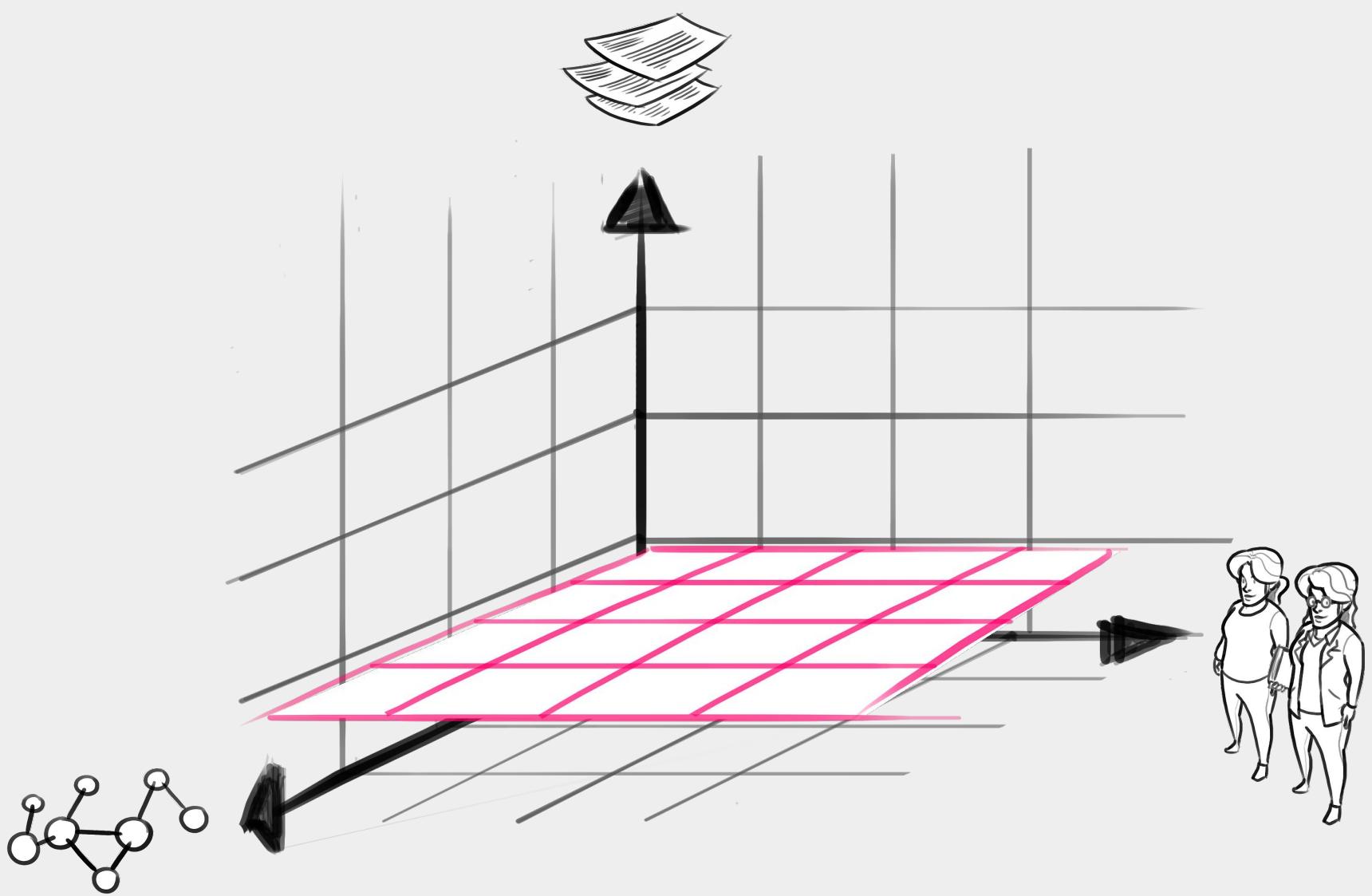
VECTOR SPACE

The final frontier



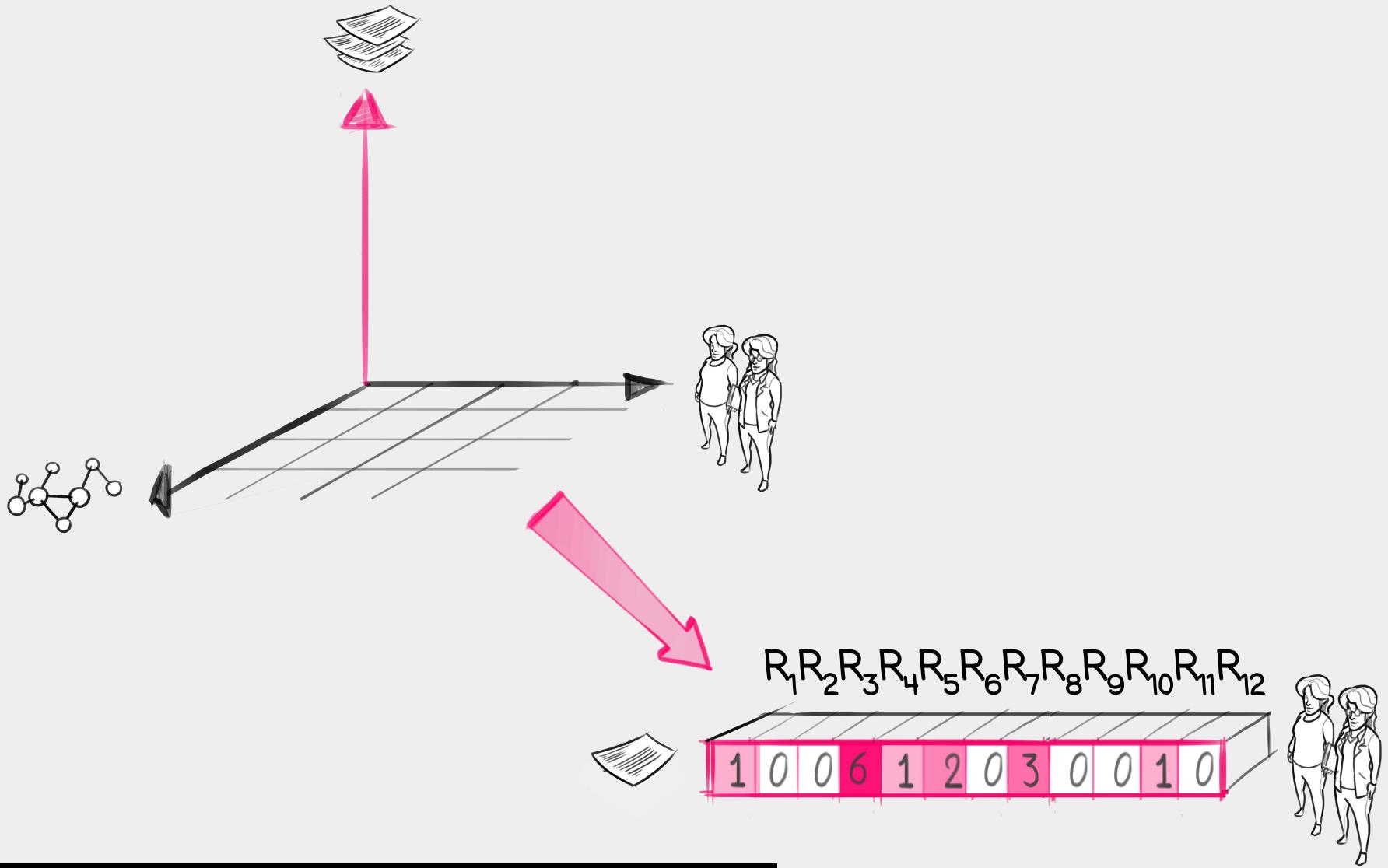
HYPER-DIMENSIONAL SPACE

Worker votes



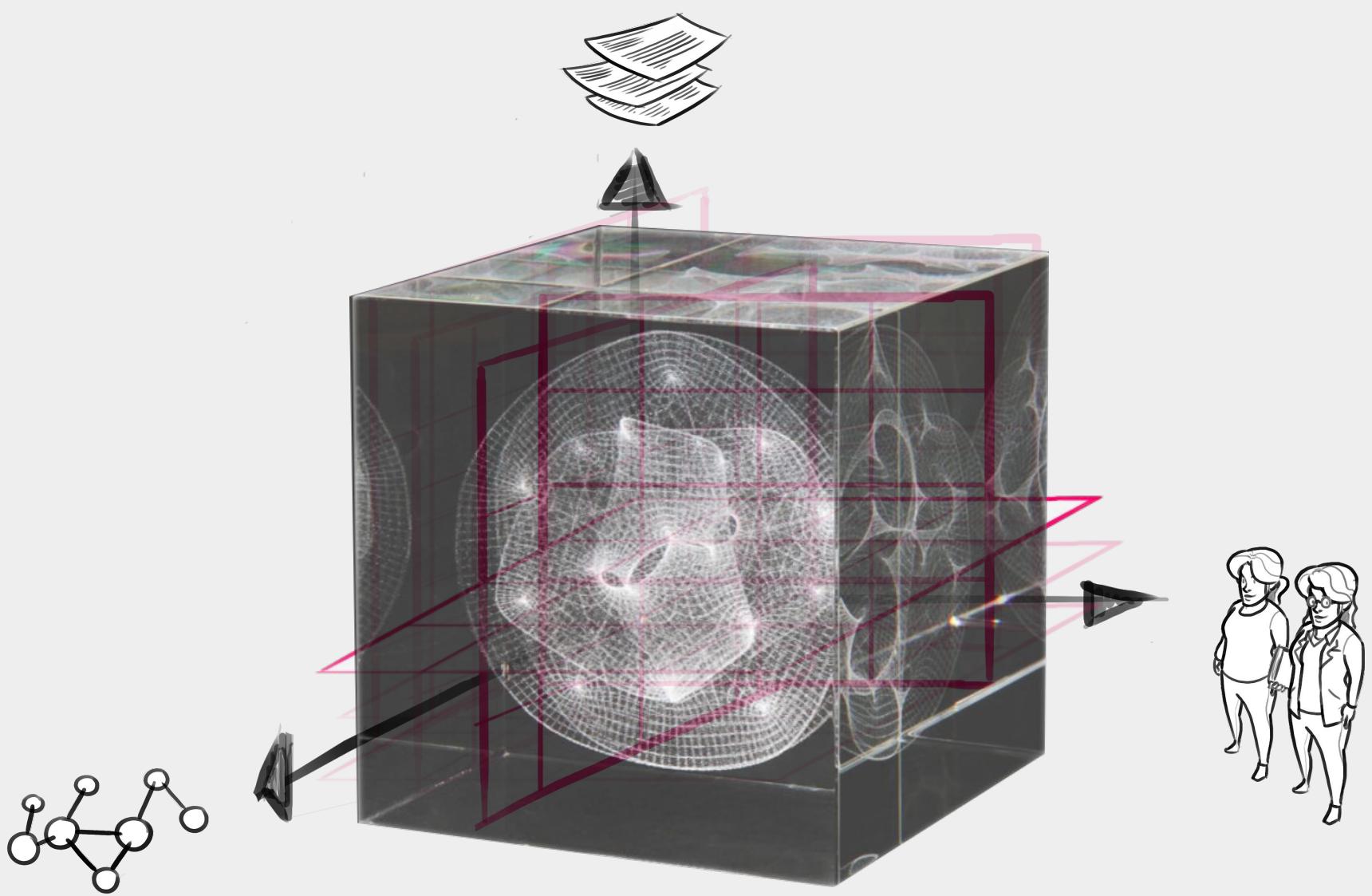
HYPER-DIMENSIONAL SPACE

Sentence Slice



HYPER-DIMENSIONAL SPACE

Sentence plane into a sentence vector



HYPER-DIMENSIONAL SPACE

3 Sentence Slices

CROWD TASK



Medical Relation Extraction

powered by
CrowdTruth 
The framework for crowdsourcing ground truth data

1

In the following sentence:

Sentence: _____

Among 56 subjects reporting to a clinic with symptoms of **malaria**, 53 (95%) had ordinarily effective levels of **chloroquine** in blood.

2

Is **chloroquine** related to **malaria**? Choose all that apply.

Treats

Diagnosed By

Causes

Location

✓ Manifestation

Contraindicates

✓ Associated With

Is A

Part Of

✓ Symptom

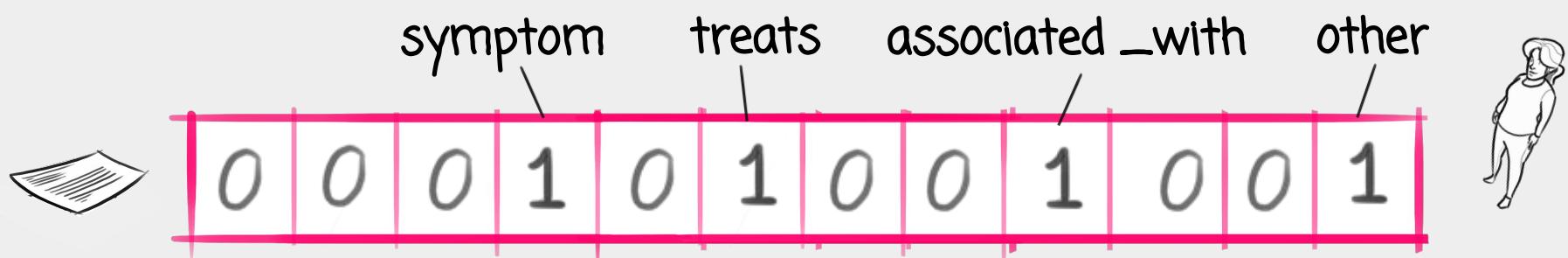
✓ Other

None

SYMPOTM: Deviation from normal function indicating the presence of disease or abnormality, e.g. pain is a symptom of a broken arm.

WORKER VECTOR FOR A SENTENCE

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



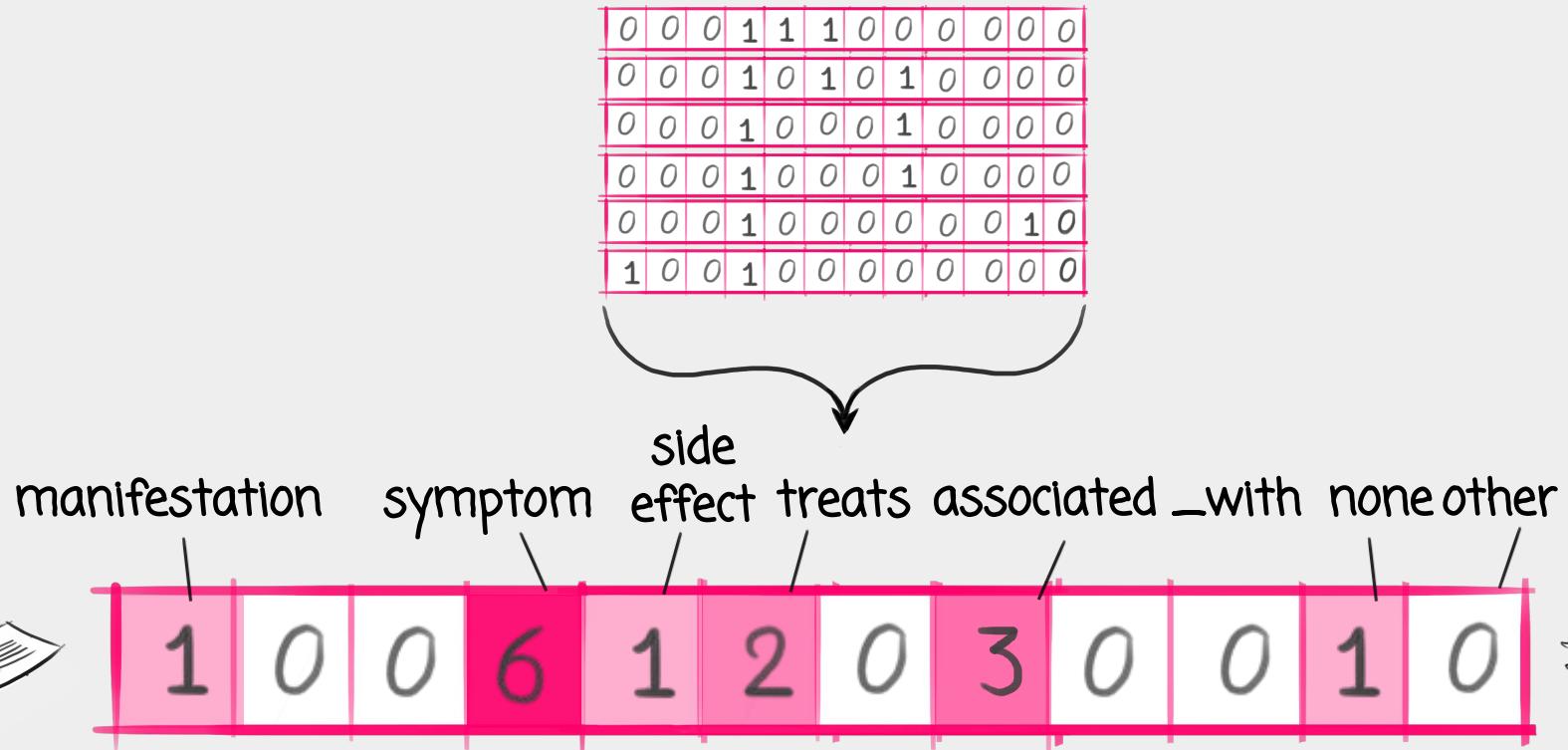
MANY WORKERS FOR THE SAME SENTENCE

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

	symptom		treats		associated _with	other					
	0	0	0	1	1	1	0	0	0	0	0
	0	0	0	1	0	1	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0	0
	0	0	0	1	0	0	0	0	0	0	1
	1	0	0	1	0	0	0	0	0	0	0

ALL WORKER VECTORS AGGREGATED IN A SENTENCE VECTOR

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



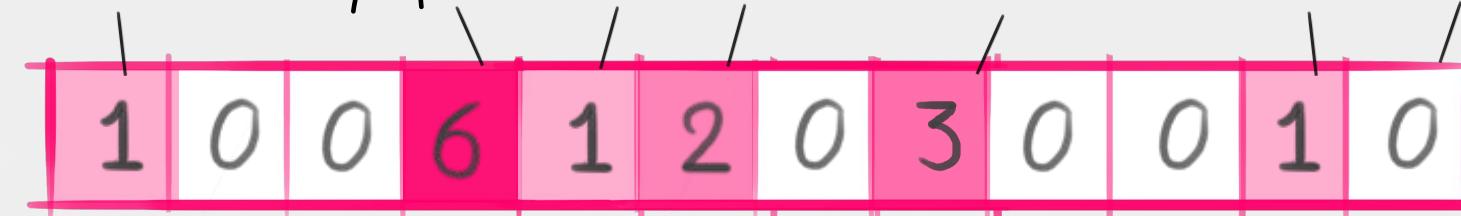
SENTENCE VECTORS FOR THE 3 SENTENCES



manifestation contraindicates treats associated_with none other



manifestation symptom effect treats associated_with none other





TIME TO DISRUPT THE COMFORT ZONE

One truth: knowledge acquisition assumes one correct interpretation for every example

All examples are created equal: one is not more important than another, they are all either true or false

Disagreement bad: when people disagree, they don't understand the problem

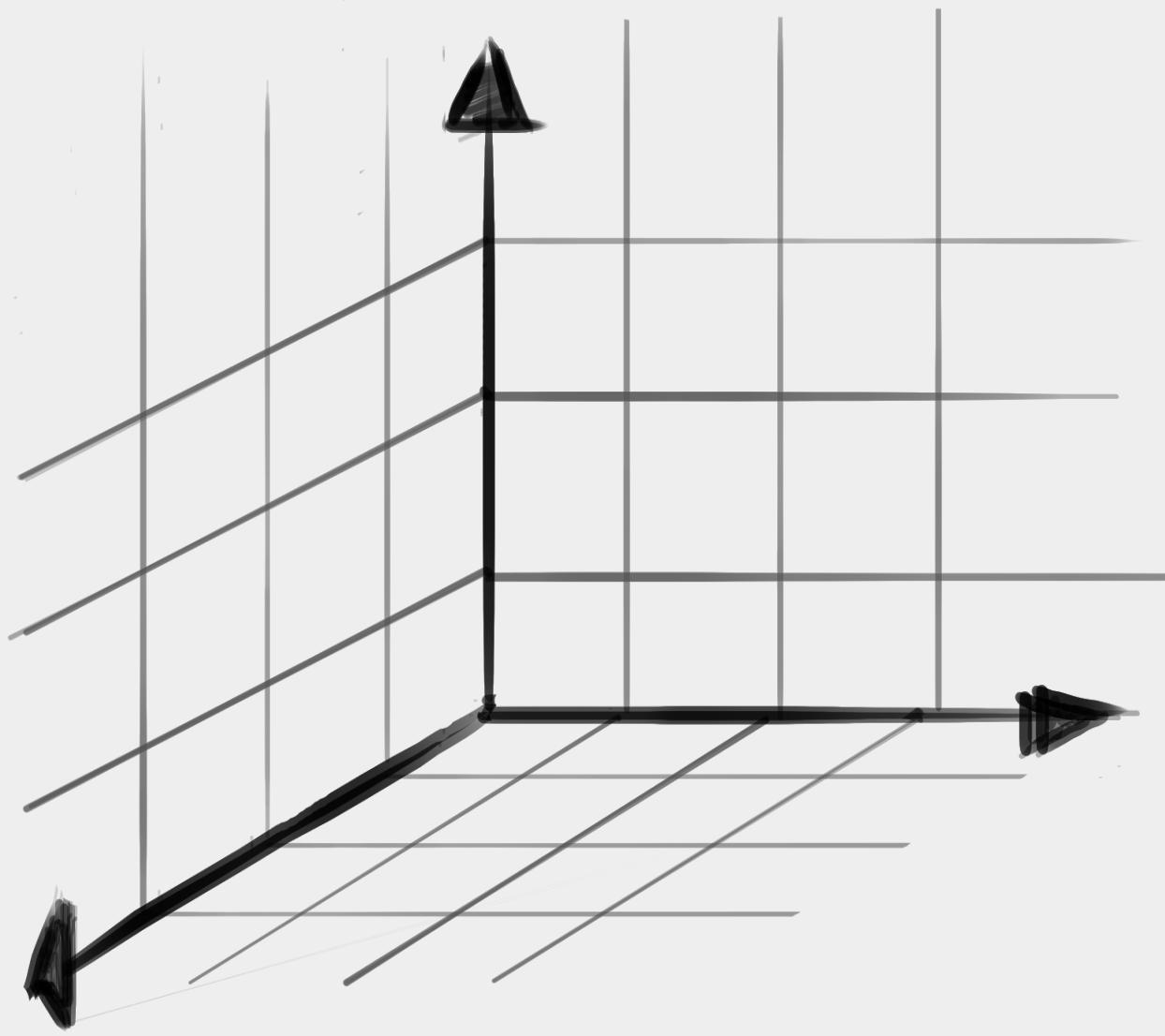
Experts rule: knowledge is captured from domain experts

One is enough: knowledge by a single expert is sufficient

Detailed explanations help: if examples cause disagreement - add instructions

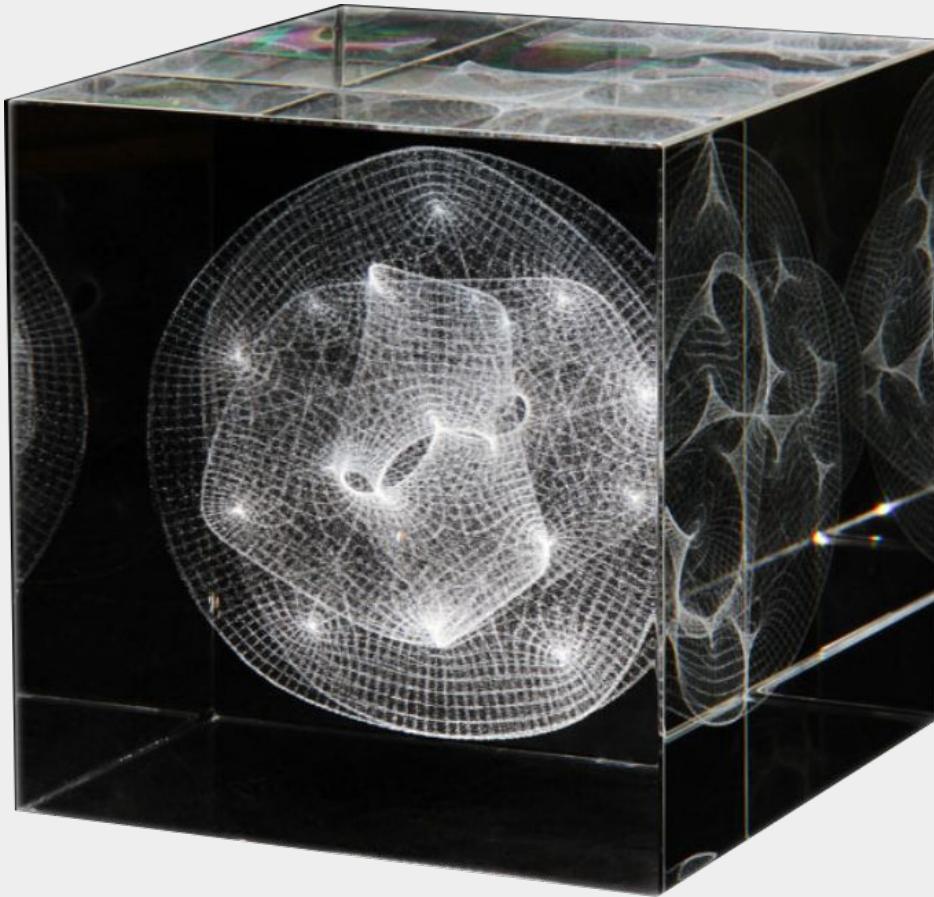
Once done, forever valid: knowledge is not updated; new data not aligned with old

"Truth is a Lie: 7 Myths about Human Annotation", *AI Magazine* 2014, L. Aroyo, C. Welty

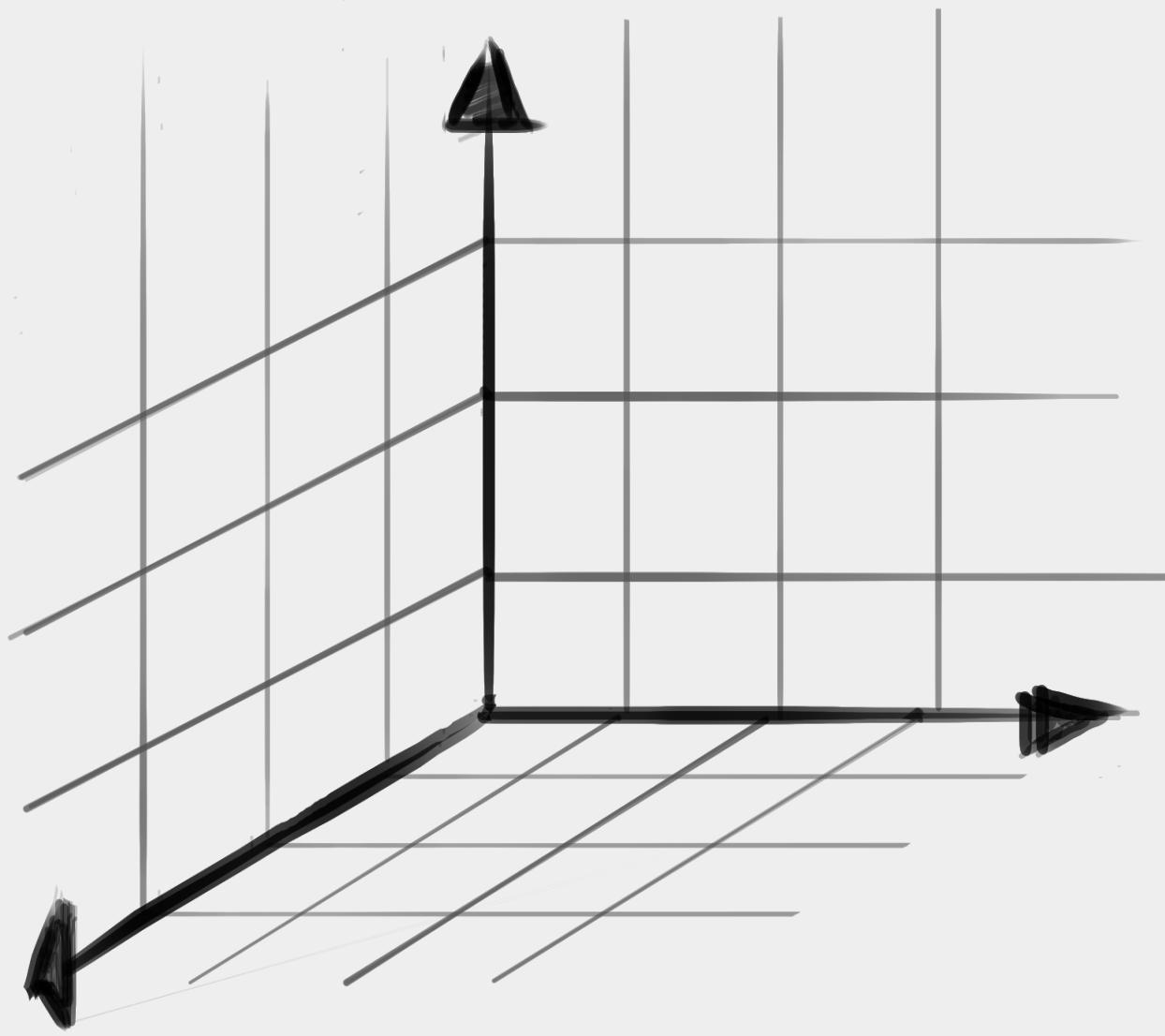


VECTOR SPACE

The final frontier

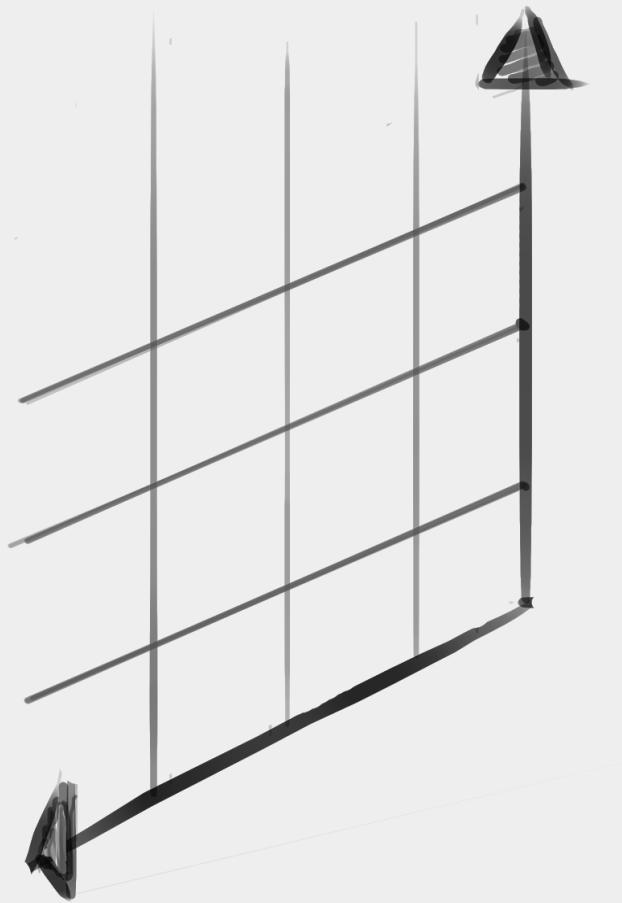


HYPER-DIMENSIONAL SPACE



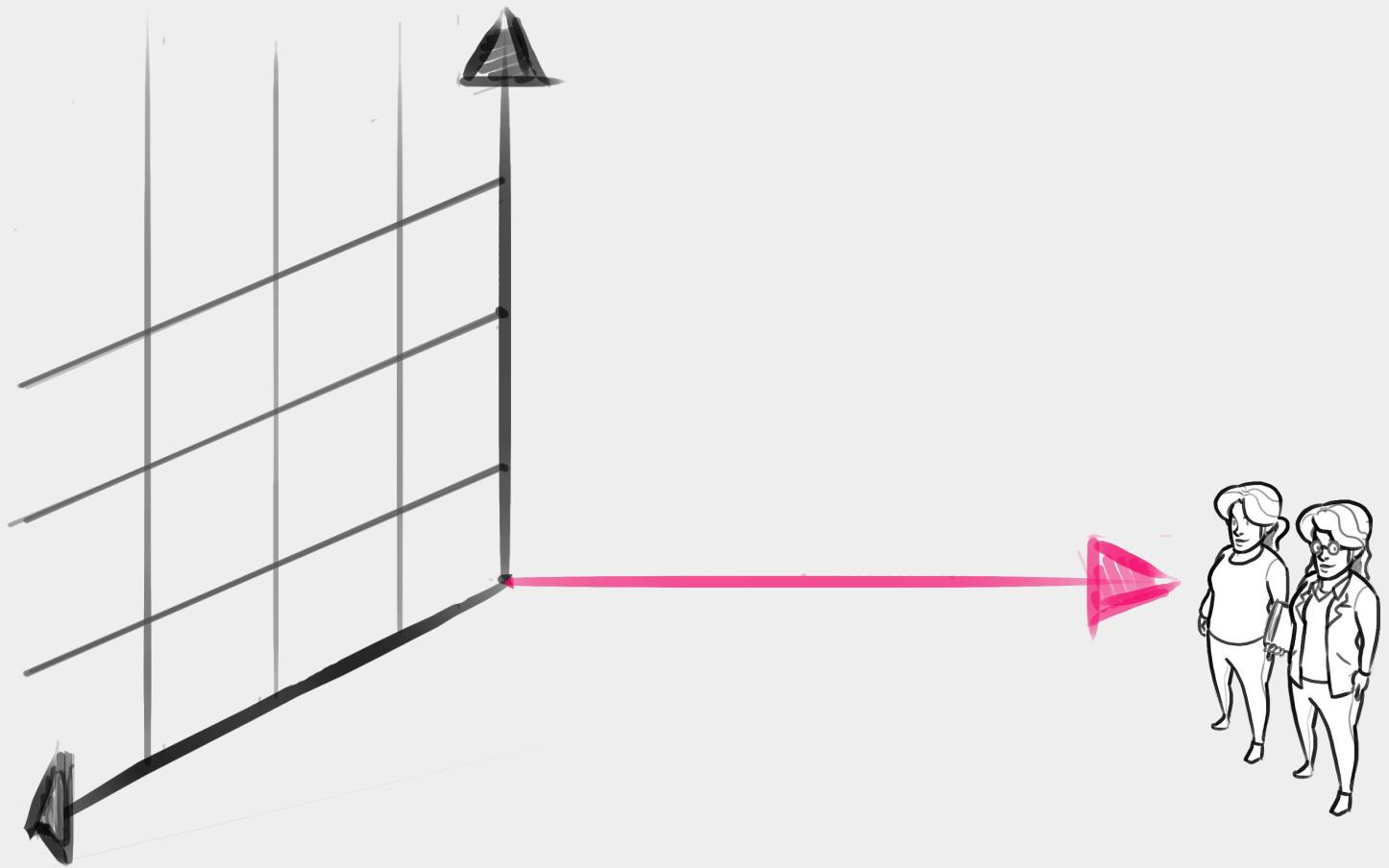
HYPER-DIMENSIONAL SPACE

3-axis tensor



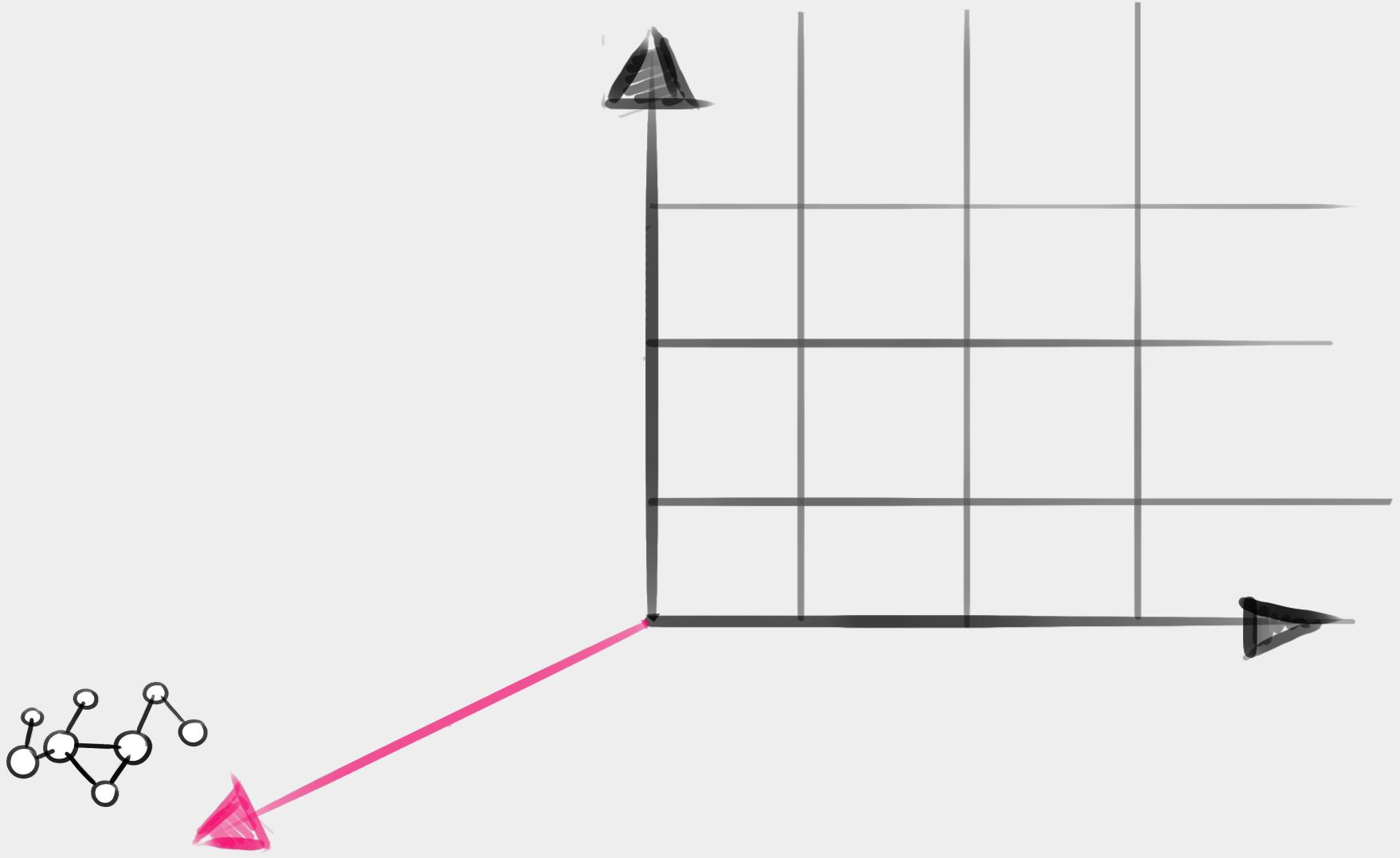
HYPER-DIMENSIONAL SPACE

... matrix



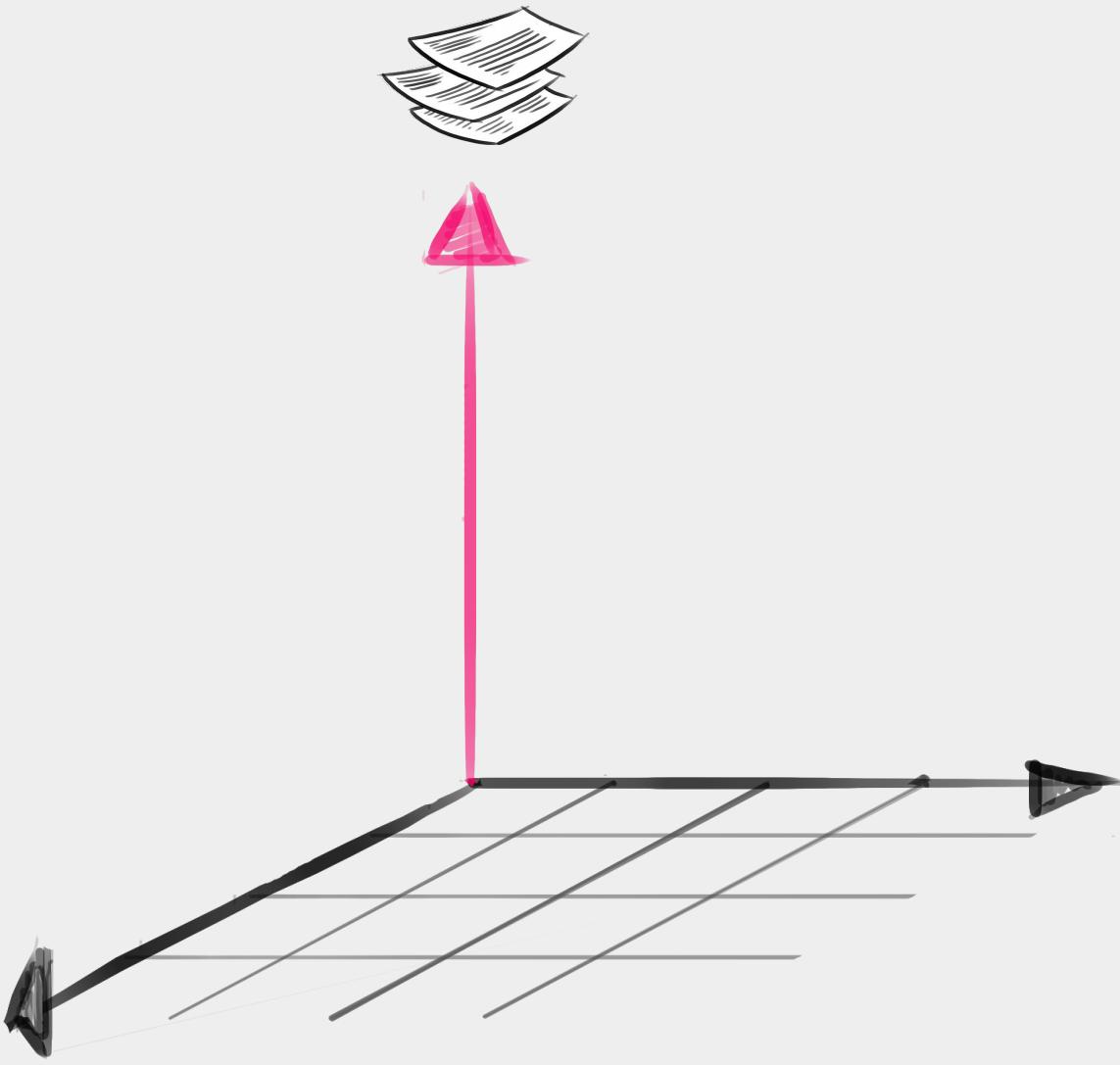
3-AXIS TENSOR

Workers axis



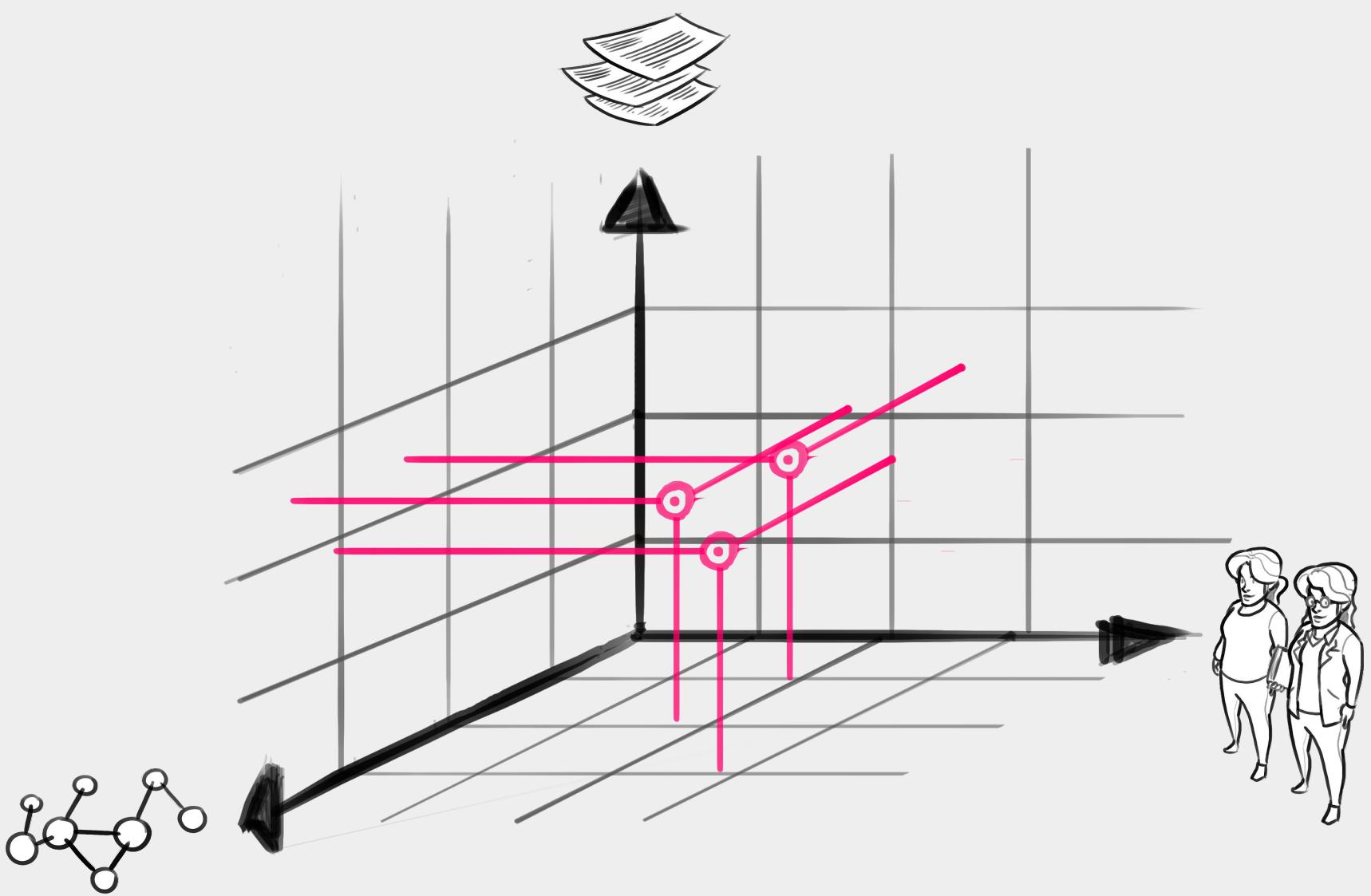
3-AXIS TENSOR

Relations axis



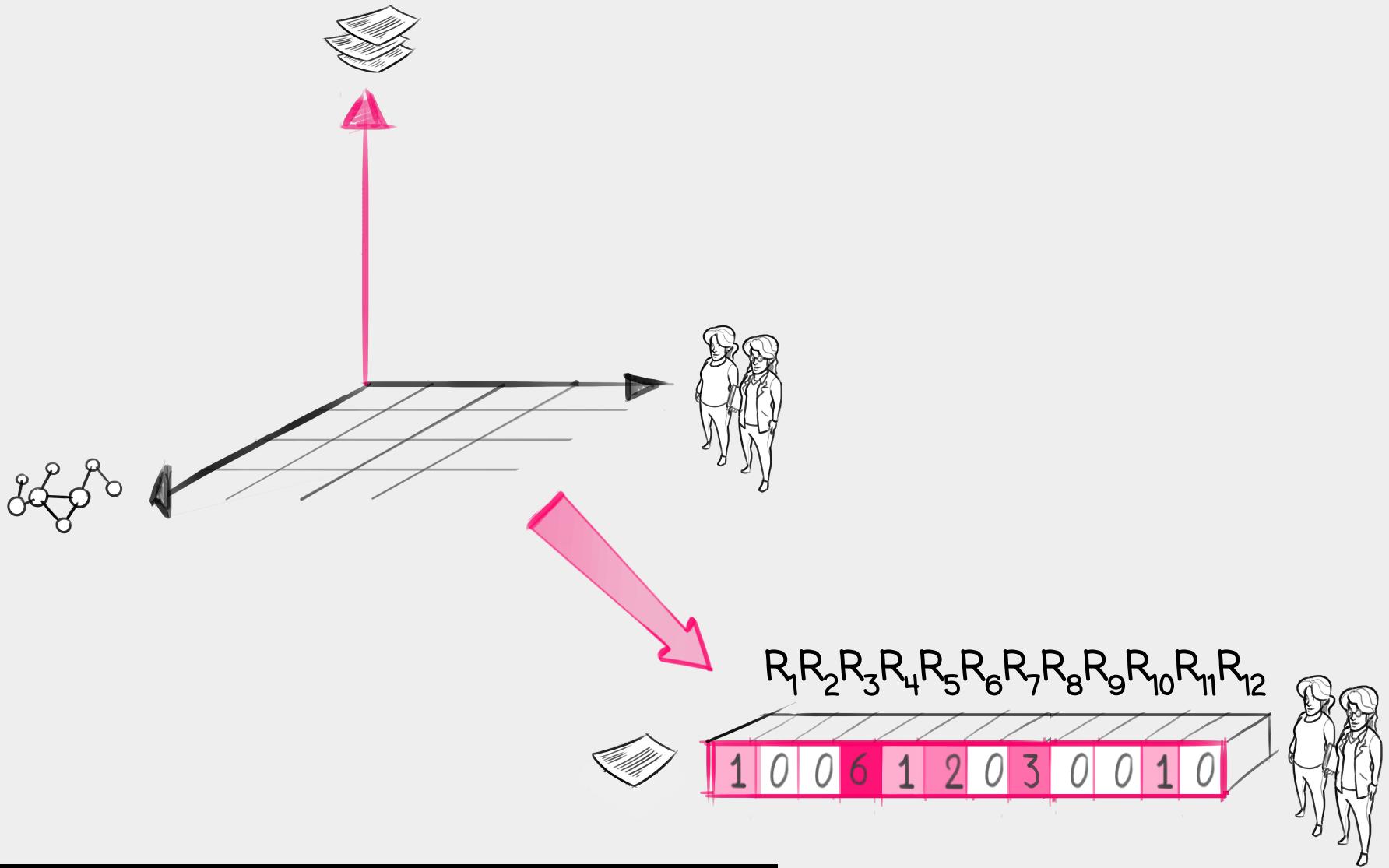
3-AXIS TENSOR

Sentences axis



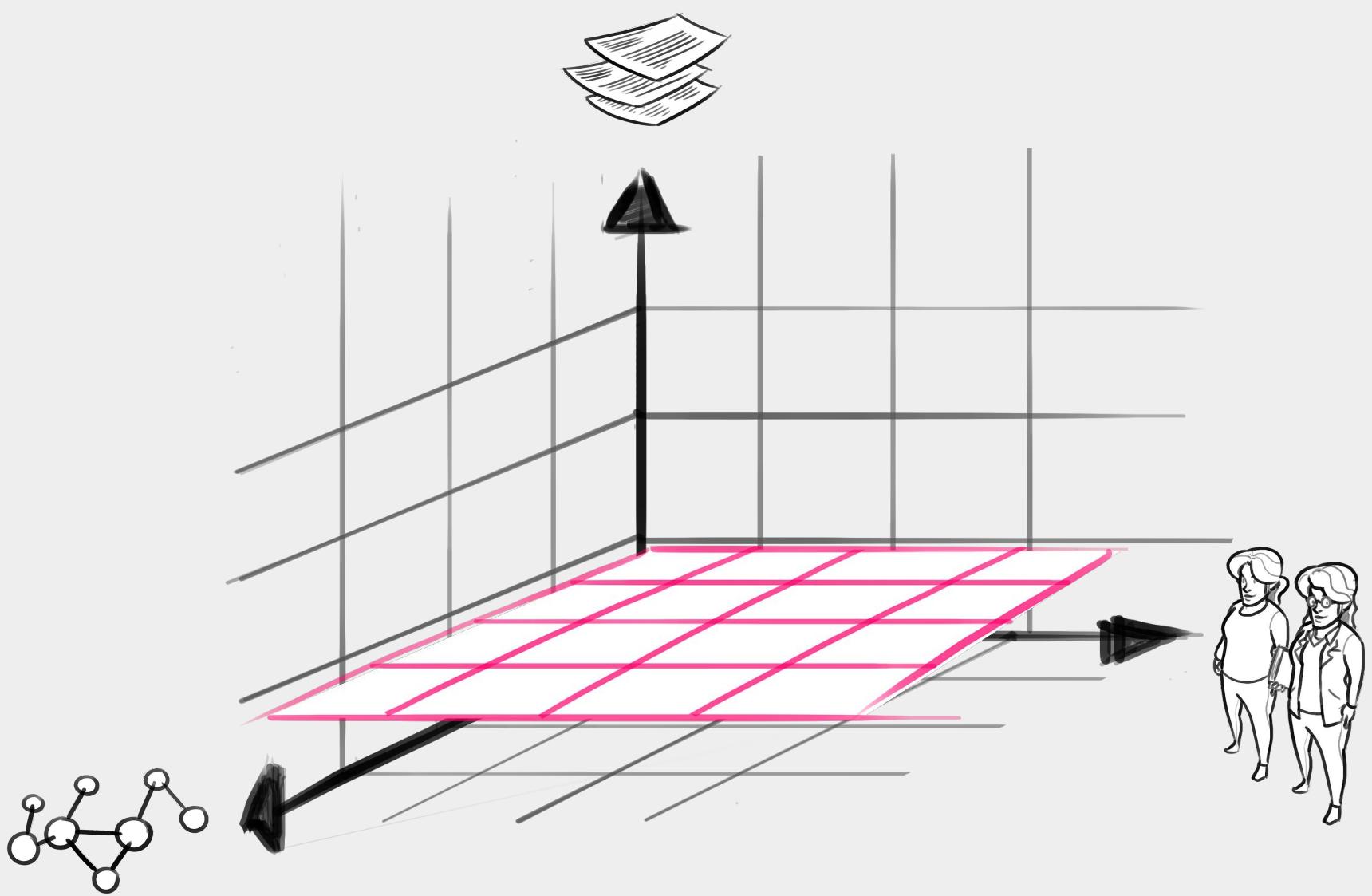
HYPER-DIMENSIONAL SPACE

Worker votes



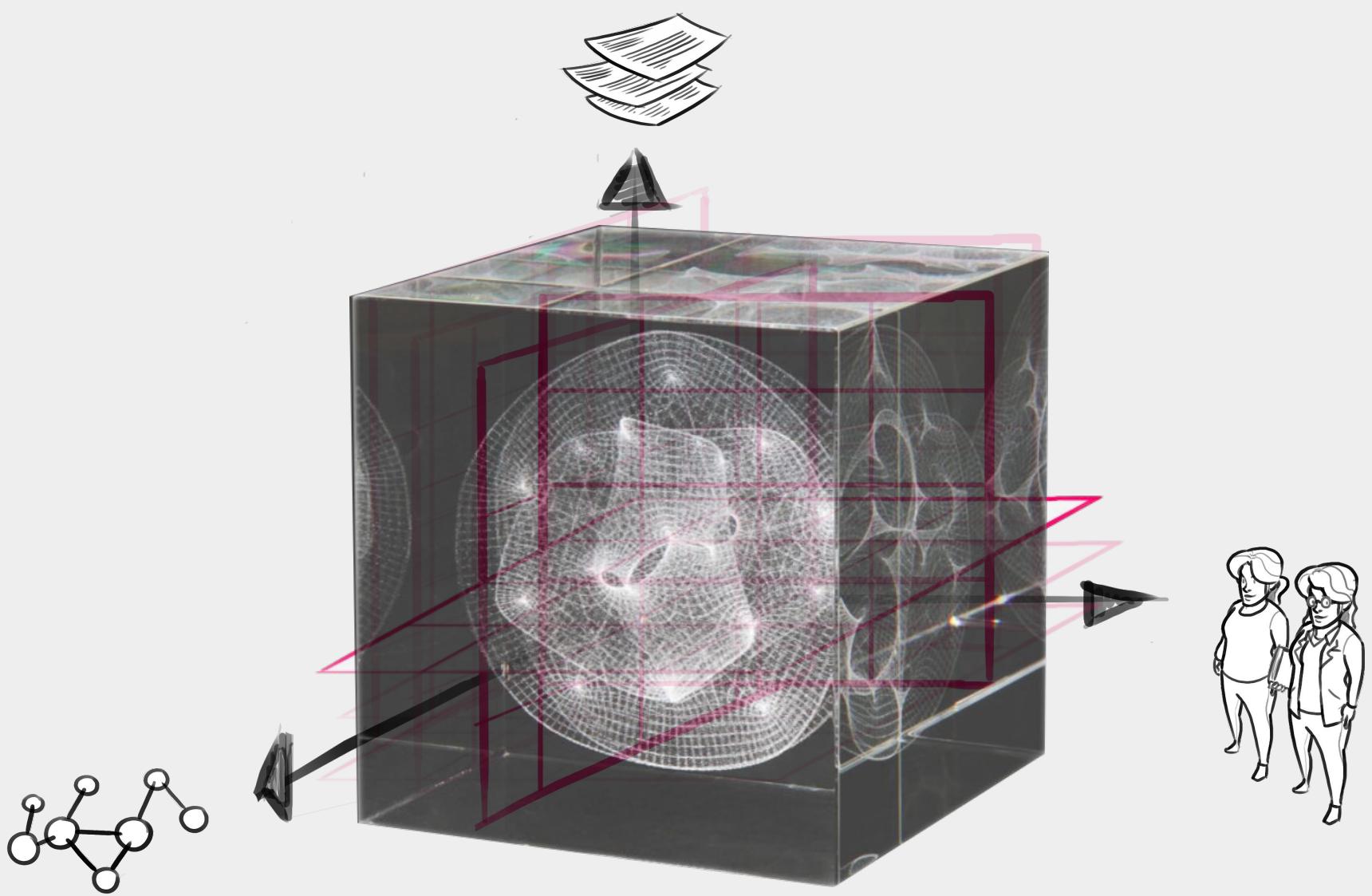
HYPER-DIMENSIONAL SPACE

Sentence plane into a sentence vector



HYPER-DIMENSIONAL SPACE

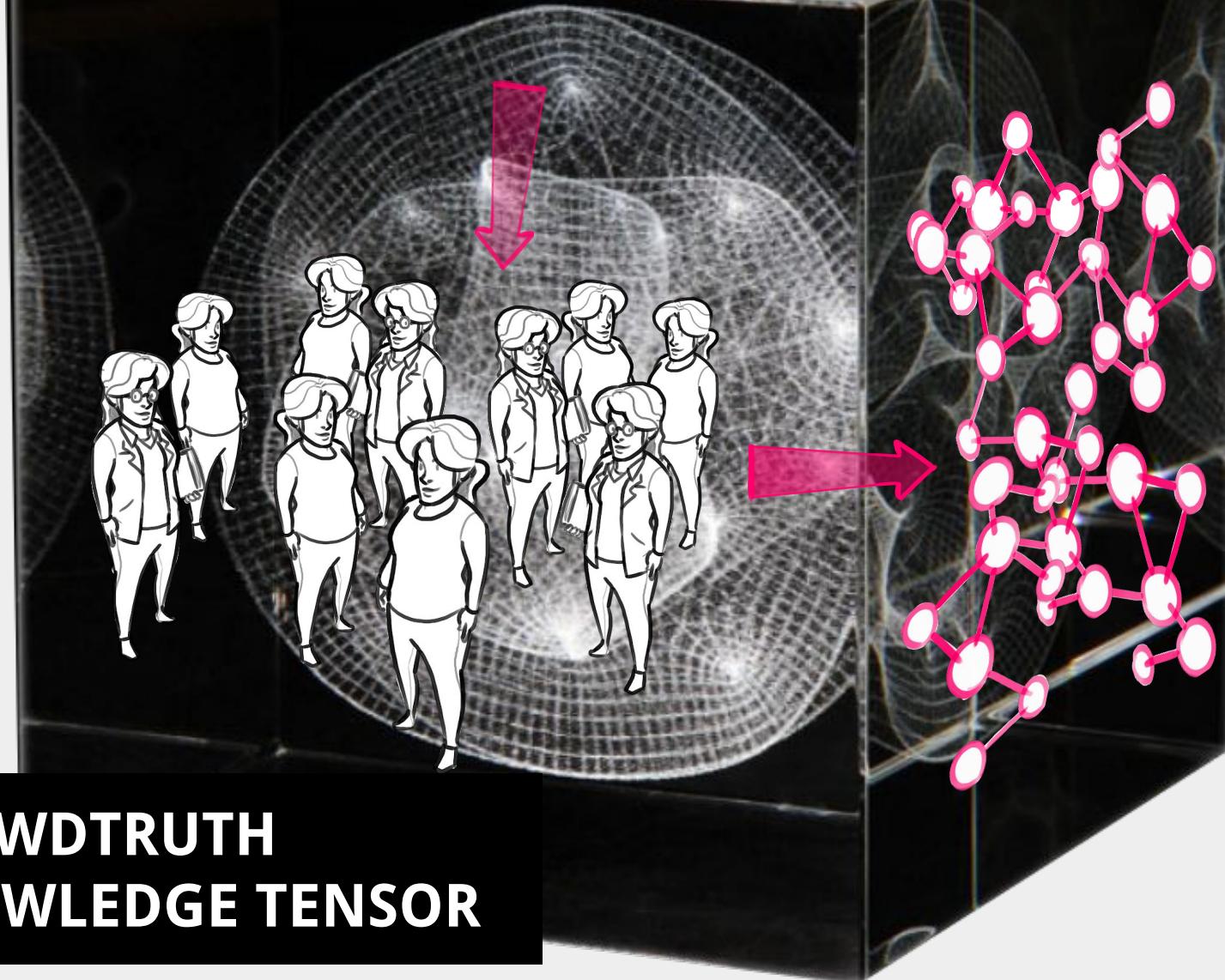
Sentence Slice



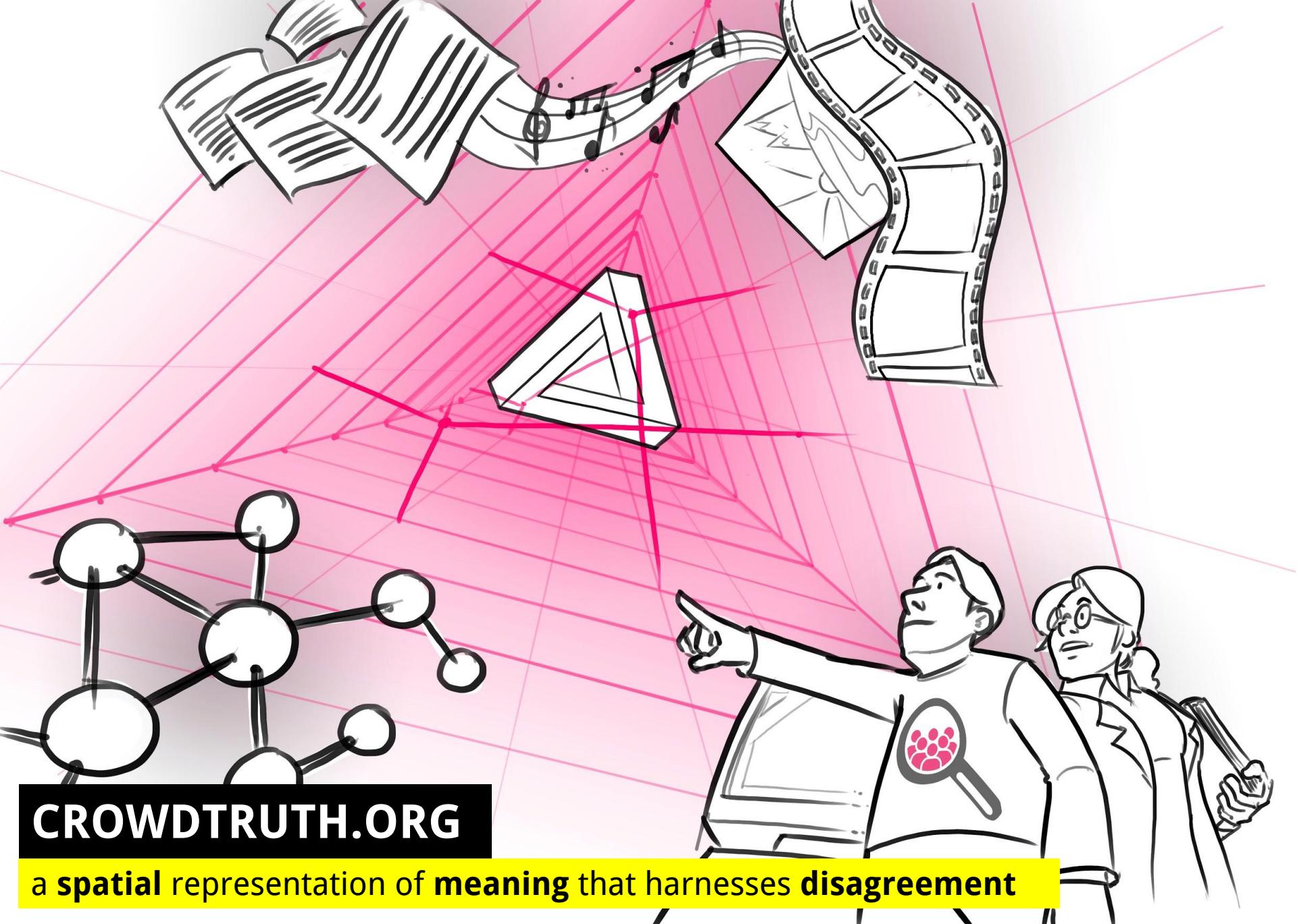
HYPER-DIMENSIONAL SPACE

3 Sentence Slices

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant *P. falciparum* **MALARIA** has not been reported.

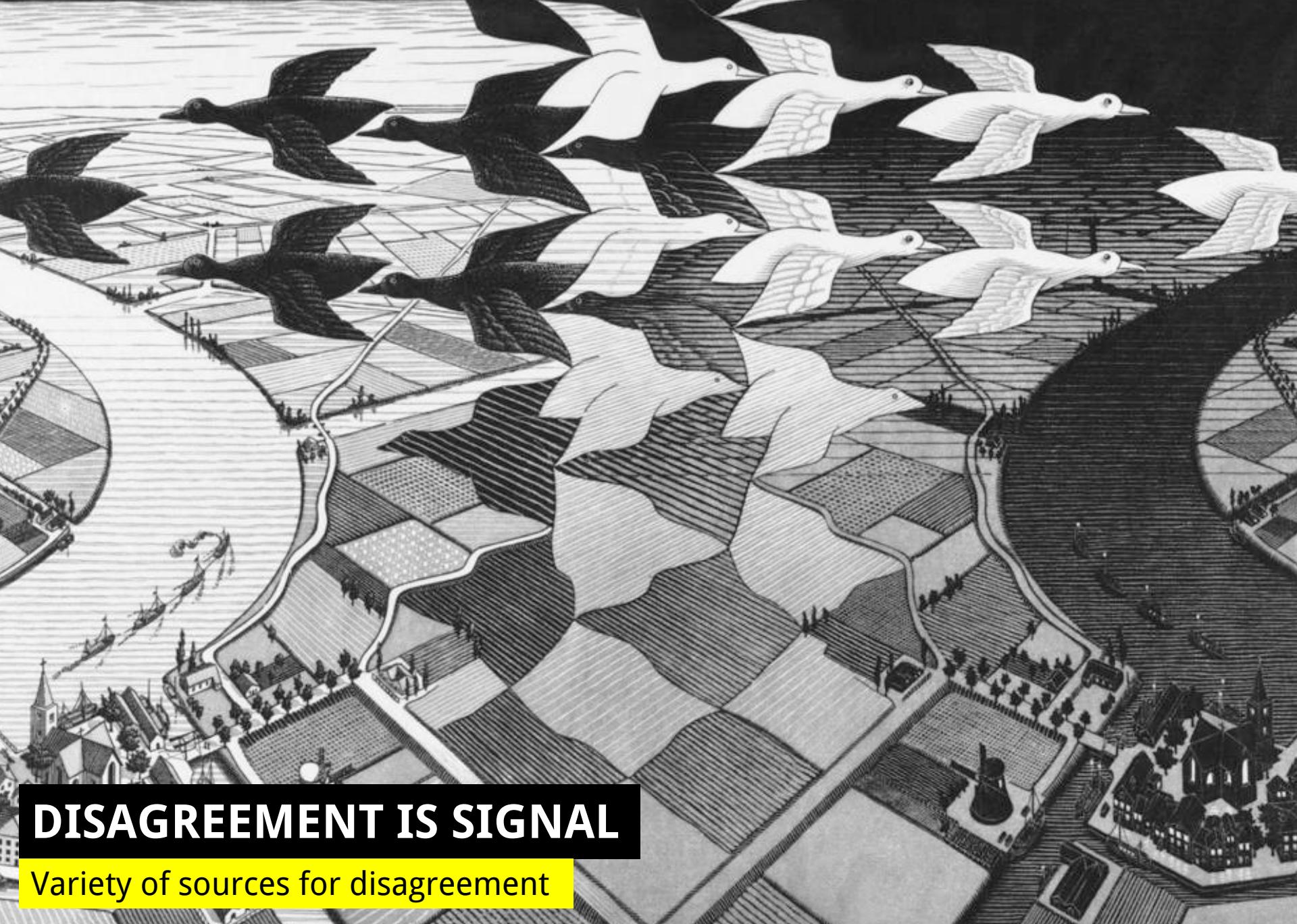


CROWDTRUTH KNOWLEDGE TENSOR



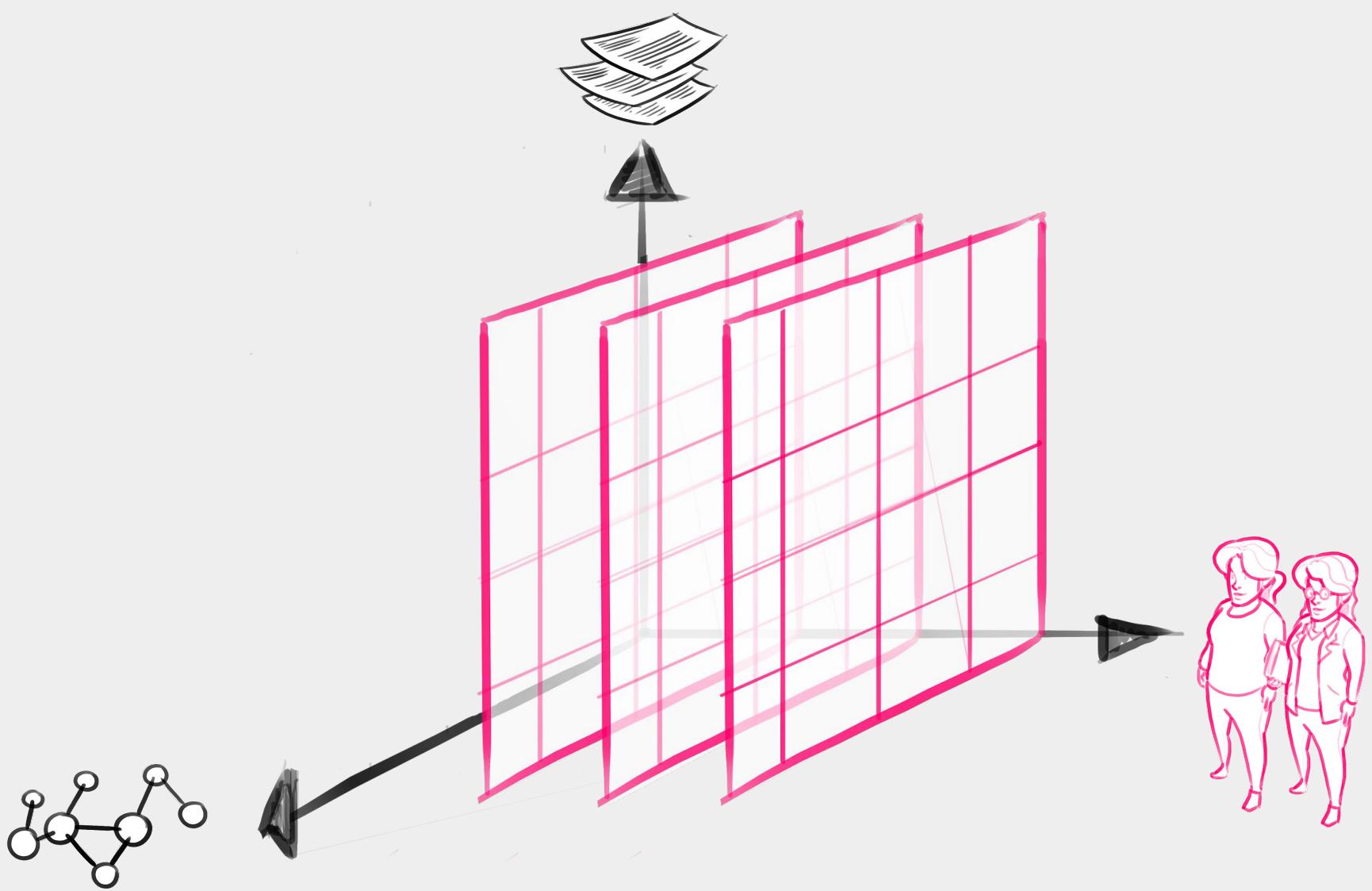
CROWDTRUTH.ORG

a spatial representation of **meaning** that harnesses **disagreement**



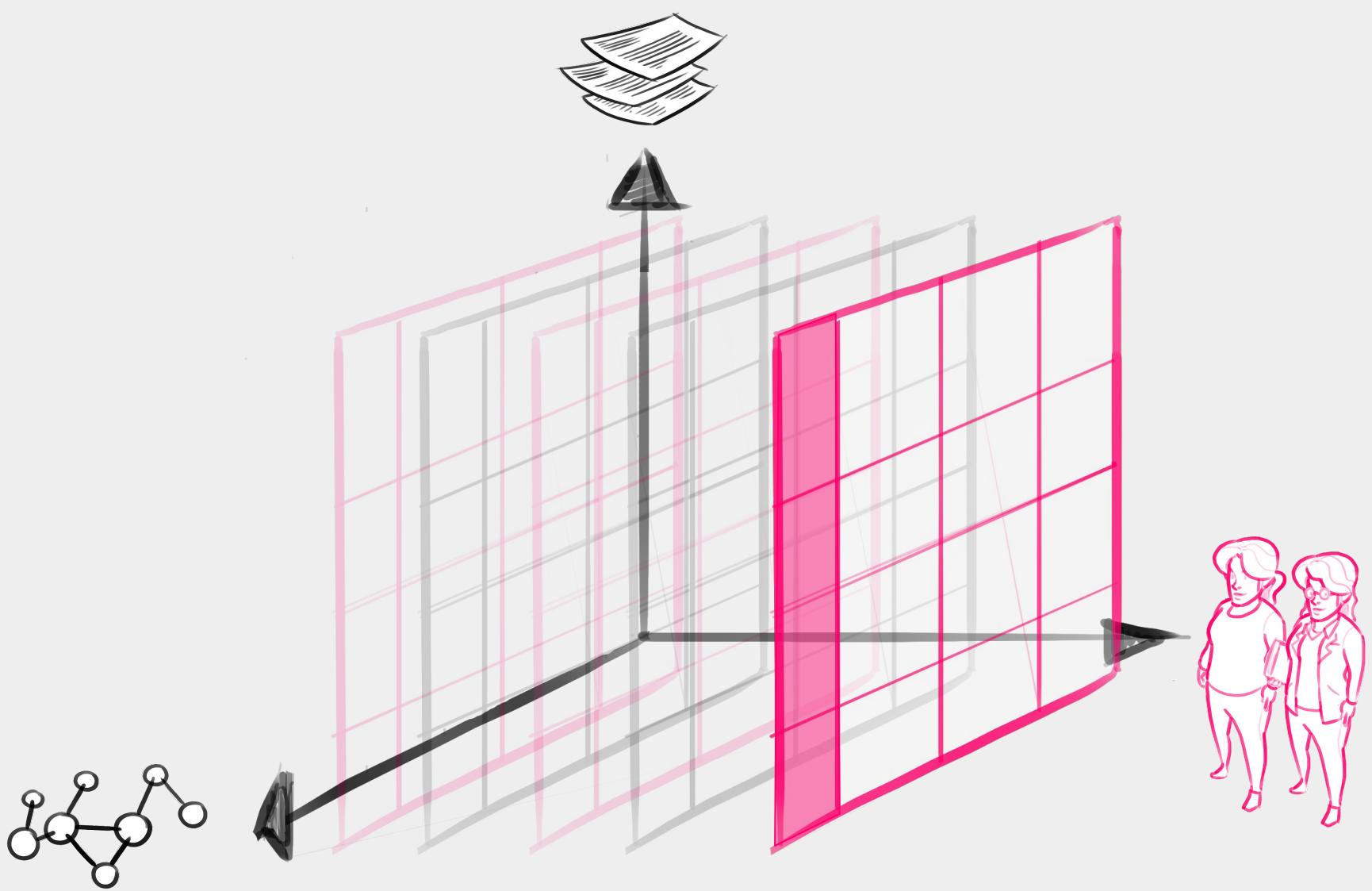
DISAGREEMENT IS SIGNAL

Variety of sources for disagreement



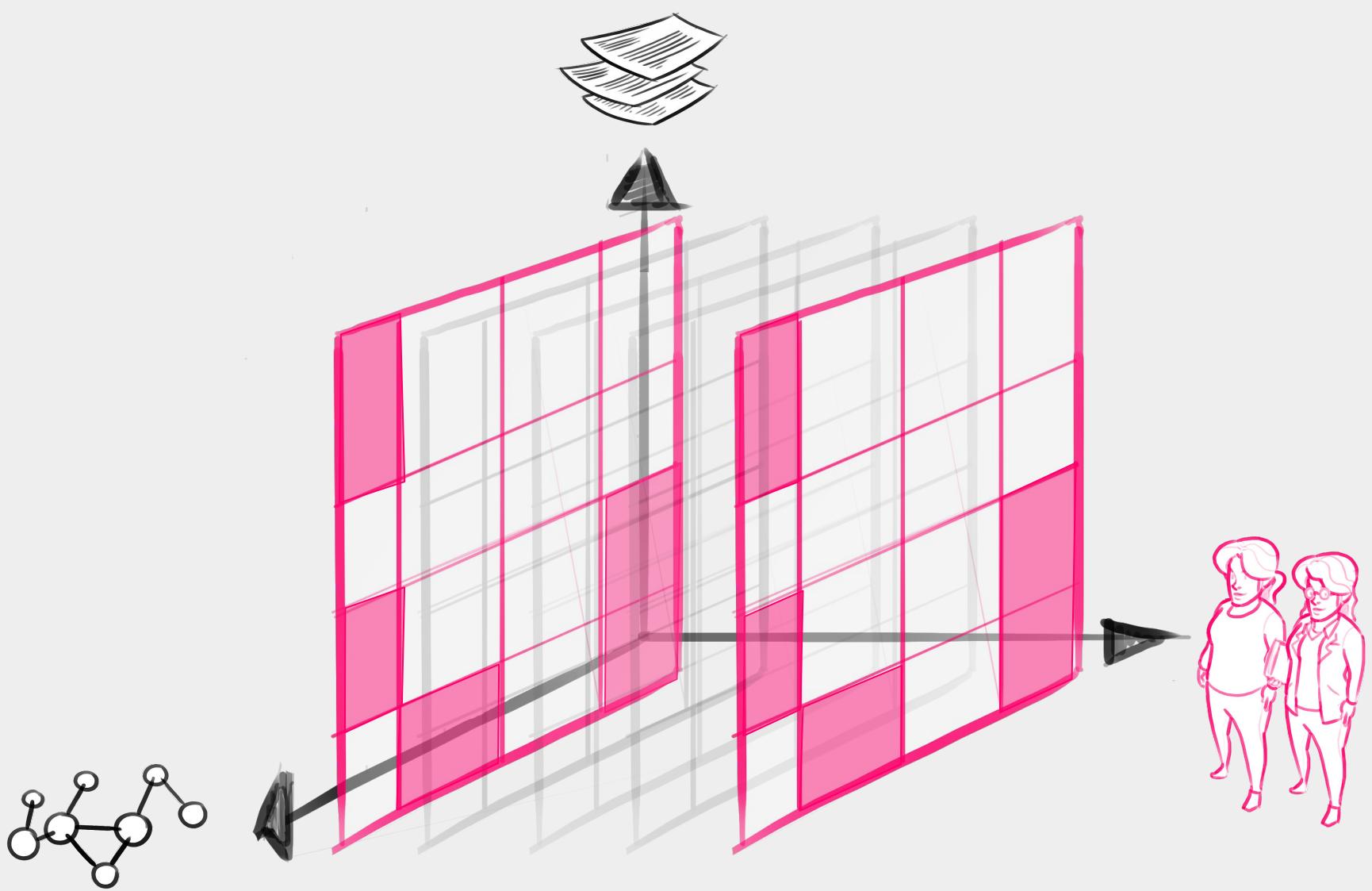
DISAGREEMENT IS SIGNAL

Source 1: People's bias & perspective



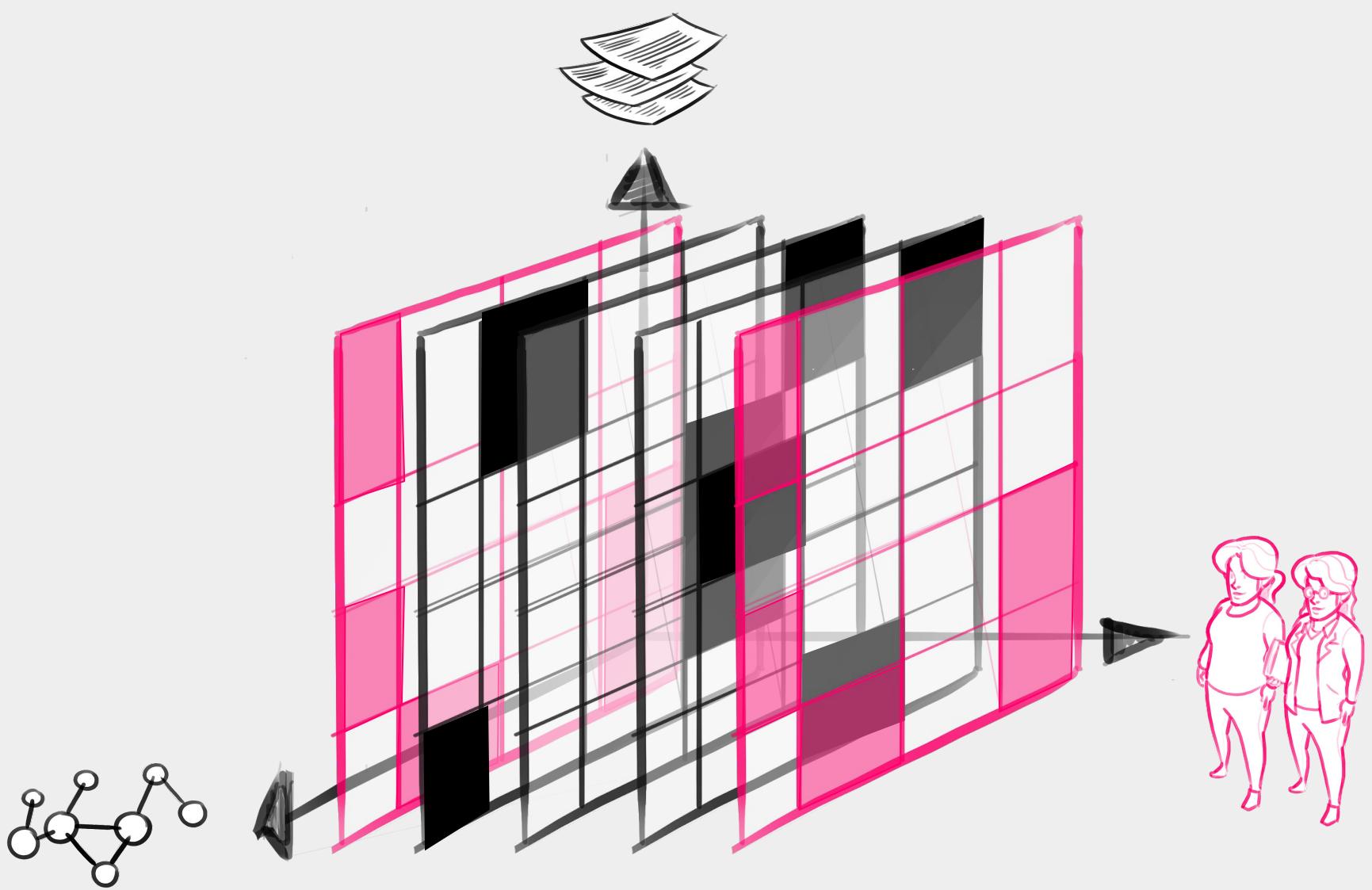
DISAGREEMENT IS SIGNAL

Source 1: Worker systematically give same answer



DISAGREEMENT IS SIGNAL

Source 1: Worker systematically give same answer



DISAGREEMENT IS SIGNAL

Source 1: Worker systematically give same answer

DO THE SENTENCES EXPRESS A TREAT RELATION?

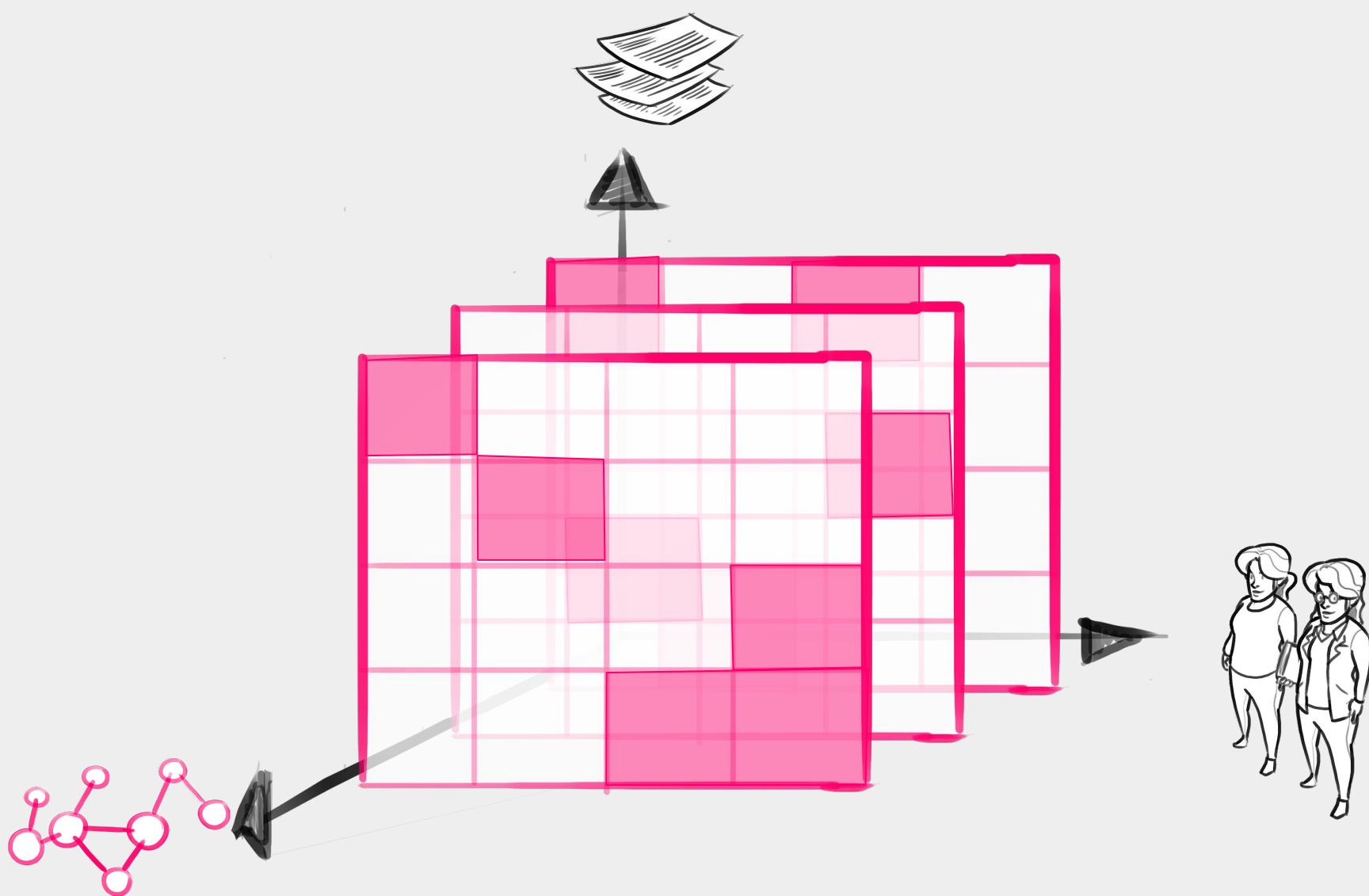
ANTIBIOTICS are the first line treatment for vectors of **TYPHUS**.

QUININE is not a reliable cure for **MALARIA**.

Worker	Sentence (1)	Sentence (2)
Worker 1	yes	no
Worker 2	yes	no
Worker 3	yes	
Worker 4		no
Worker 5	no	yes

DISAGREEMENT CAUSE

Source 1: Low Quality Workers



DISAGREEMENT IS SIGNAL

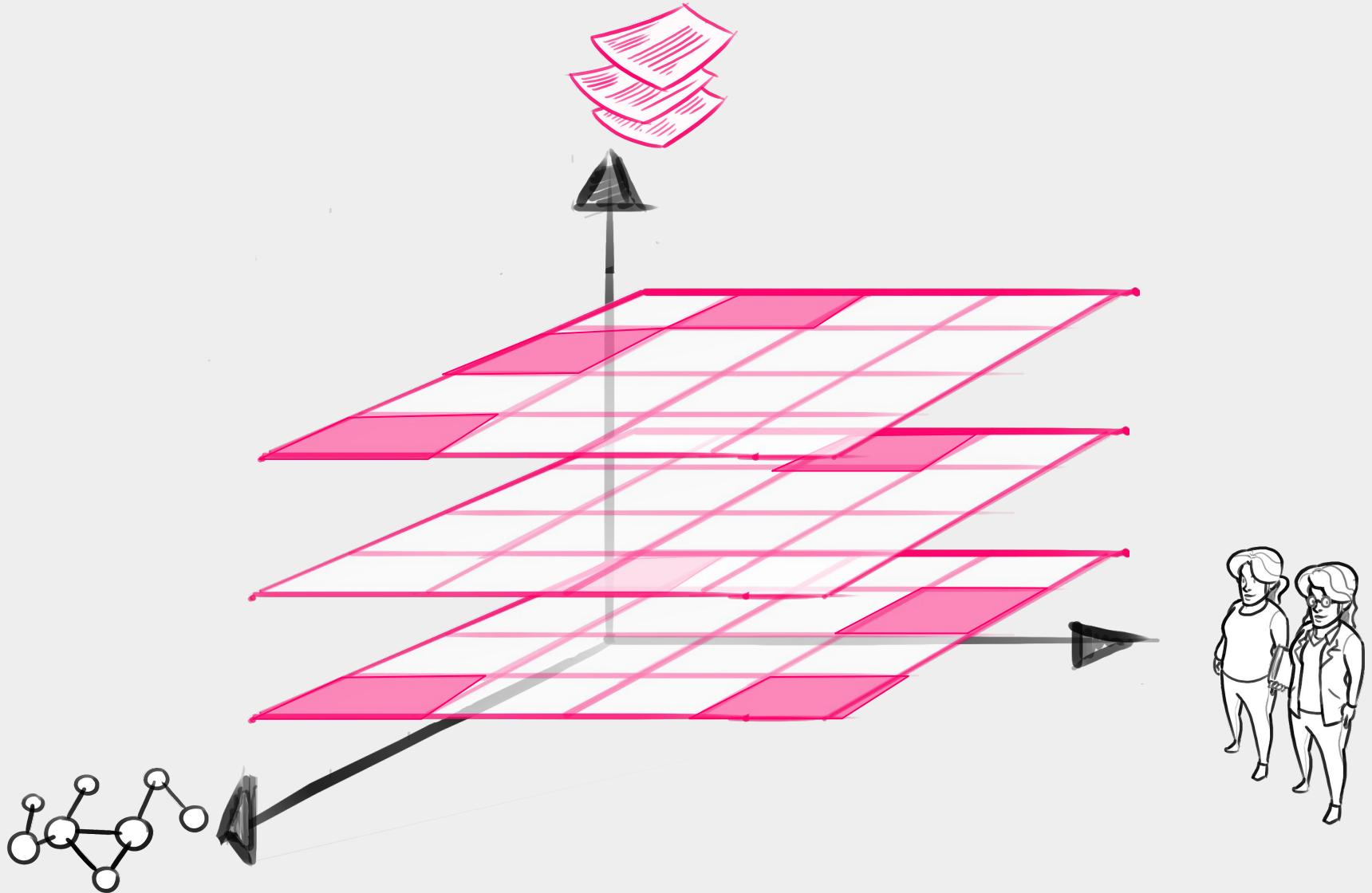
Source 2: Target semantics

WHAT IS THE RELATION EXPRESSED? TREAT OR PREVENT?

METHYLERGOMETRINE is a blood vessel constrictor most commonly used to control excessive **BLEEDING**.

DISAGREEMENT CAUSE

Source 2: Target semantics



DISAGREEMENT IS SIGNAL

Source 3: Sentence Clarity

DO THE SENTENCES EXPRESS A TREAT RELATION? BETWEEN CHLOROQUINE AND MALARIA ?

Rheumatoid arthritis and **MALARIA** have been treated with **CHLOROQUINE** for decades.

95%

For prevention of malaria, use only in individuals traveling to malarious areas where **CHLOROQUINE** resistant P. falciparum **MALARIA** has not been reported.

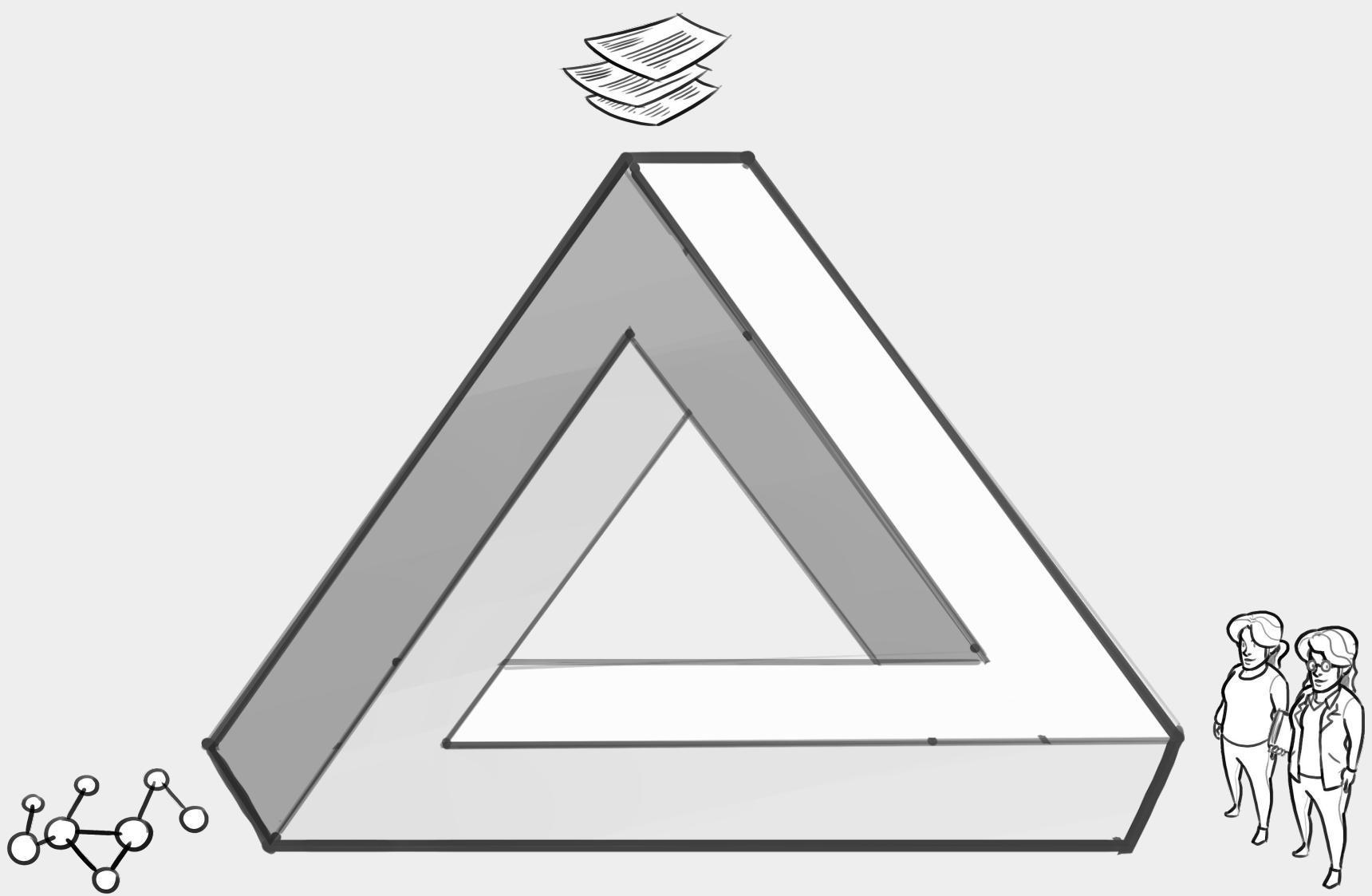
75%

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.

50%

DISAGREEMENT CAUSE

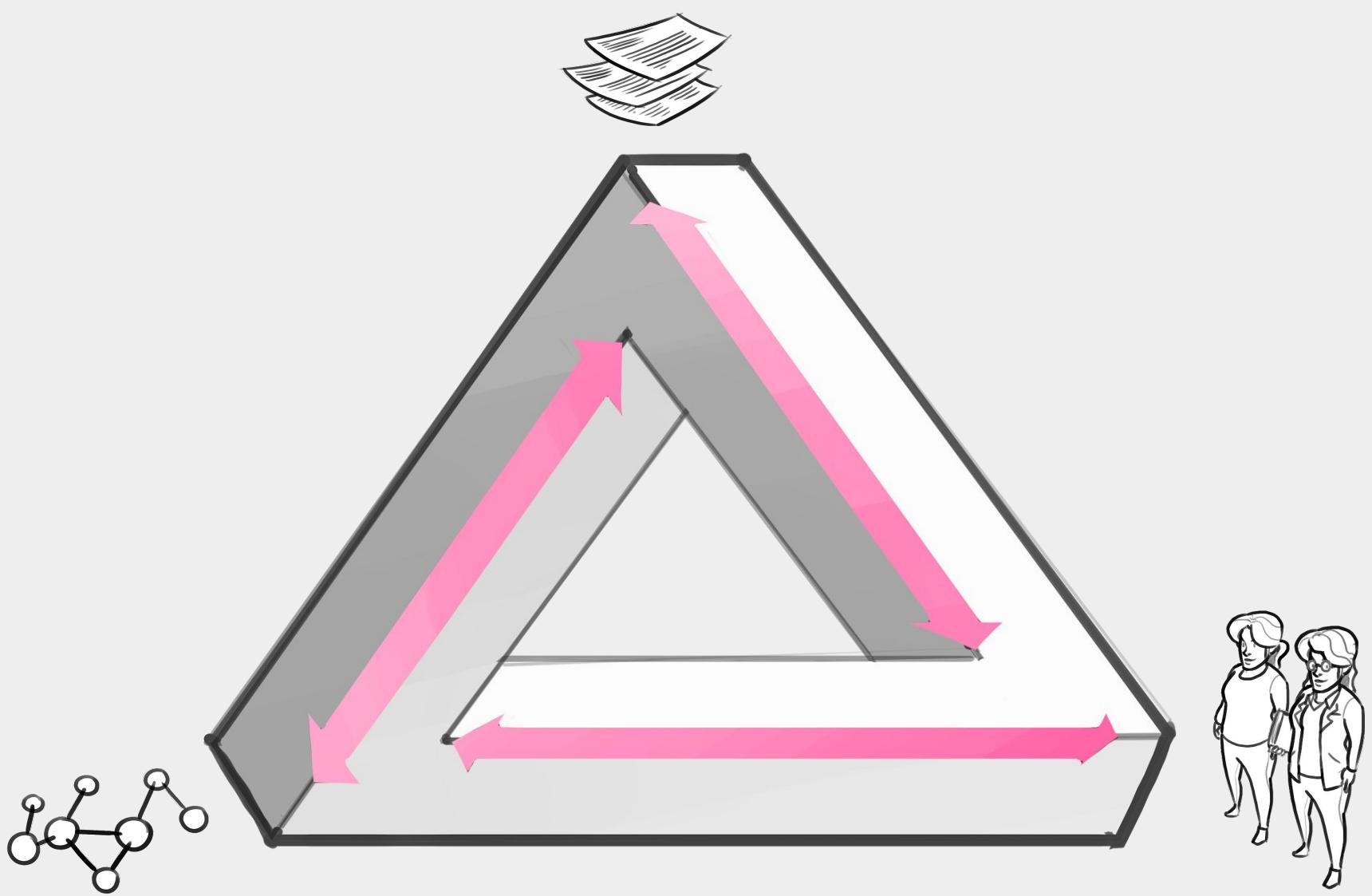
Source 3: Sentence Clarity



TRIANGLE OF MEANING

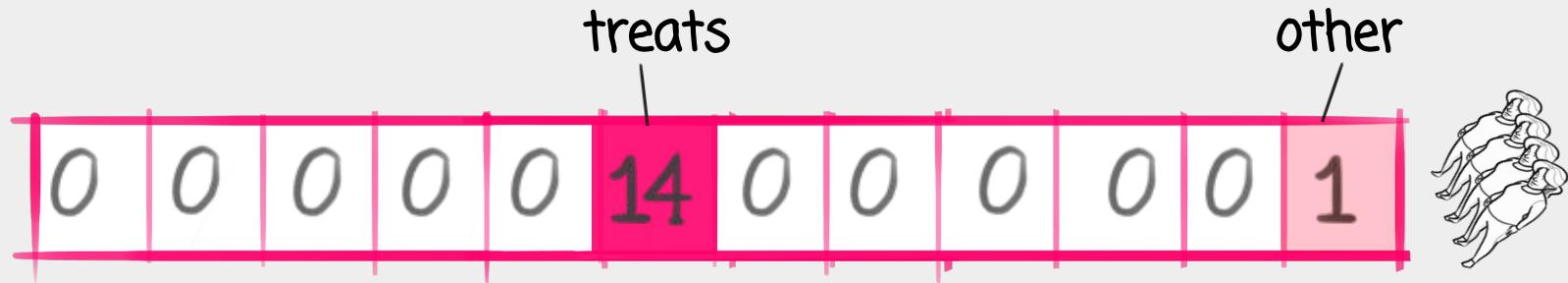
Model of semantic interpretation

*(Ogden & Richards, 1936)
The Meaning of Meaning*



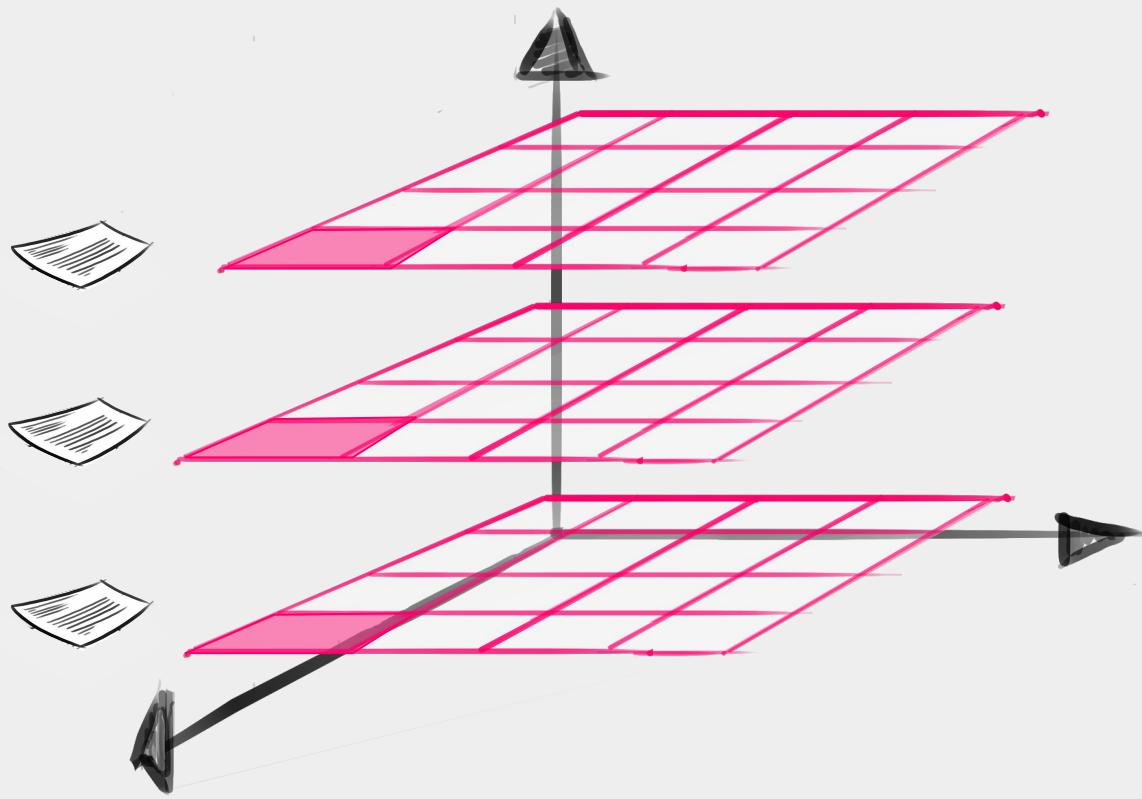
TRIANGLE OF MEANING

Model of semantic interpretation



TRIANGLE OF MEANING

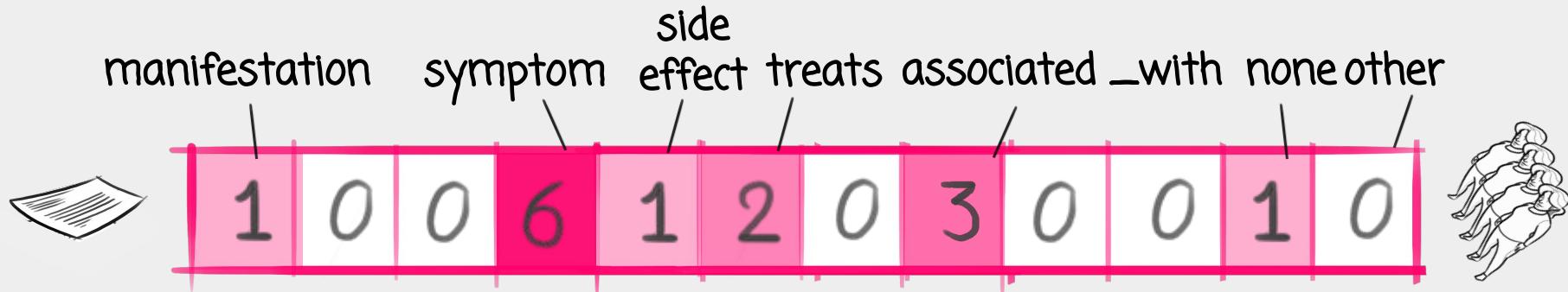
CrowdTruth metrics for quality assessment



QUALITY ASSESSMENT

Spam

Among 56 subjects reporting to a clinic with symptoms of **MALARIA** 53 (95%) had ordinarily effective levels of **CHLOROQUINE** in blood.



QUALITY ASSESSMENT

Sentence ambiguity

manifestation contraindicates treats associated_with none other



1	0	3	0	0	6	0	0	2	0	1	4
---	---	---	---	---	---	---	---	---	---	---	---



manifestation contraindicates treats associated_with none other



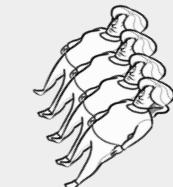
2	0	2	0	0	8	0	0	0	0	1	0
---	---	---	---	---	---	---	---	---	---	---	---



treats



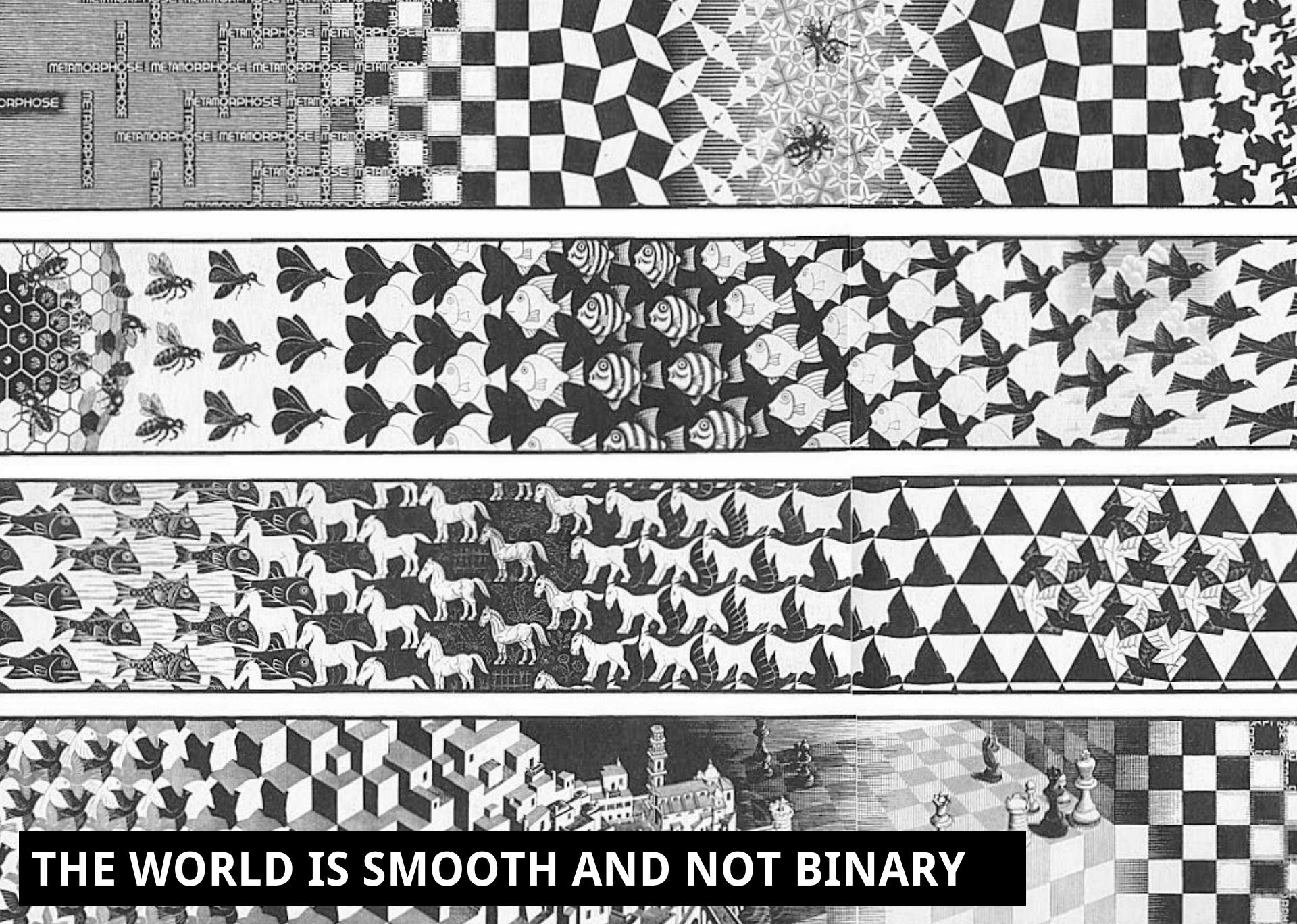
0	0	0	0	0	14	0	0	0	0	0	1
---	---	---	---	---	----	---	---	---	---	---	---



other

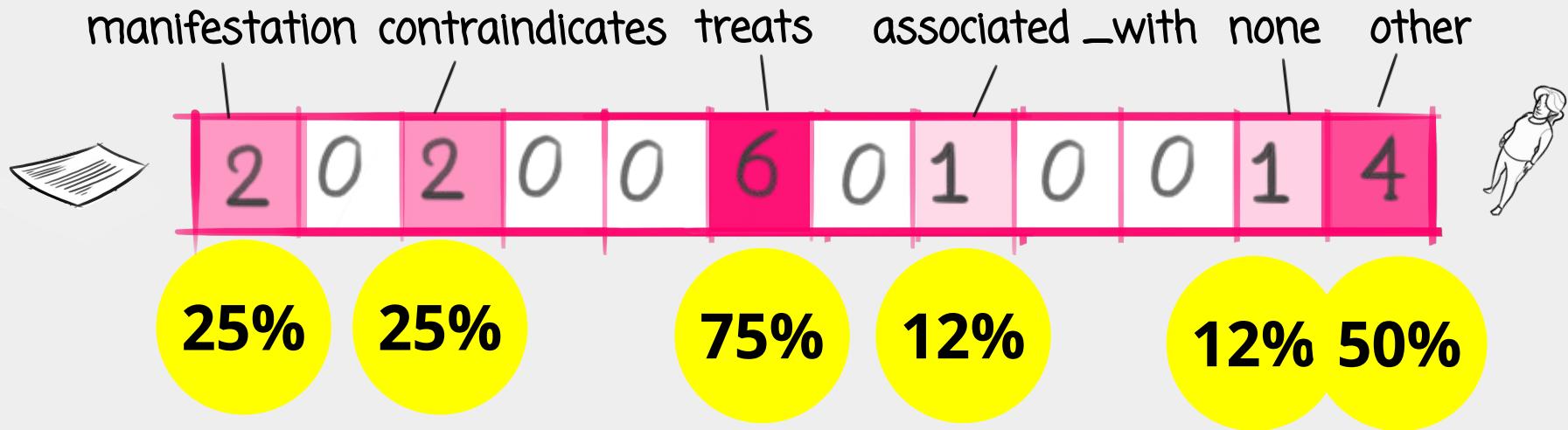
TREATS RELATION

Yes or No?



THE WORLD IS SMOOTH AND NOT BINARY

For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.



CROWDTRUTH METRICS

Agreement as percentage

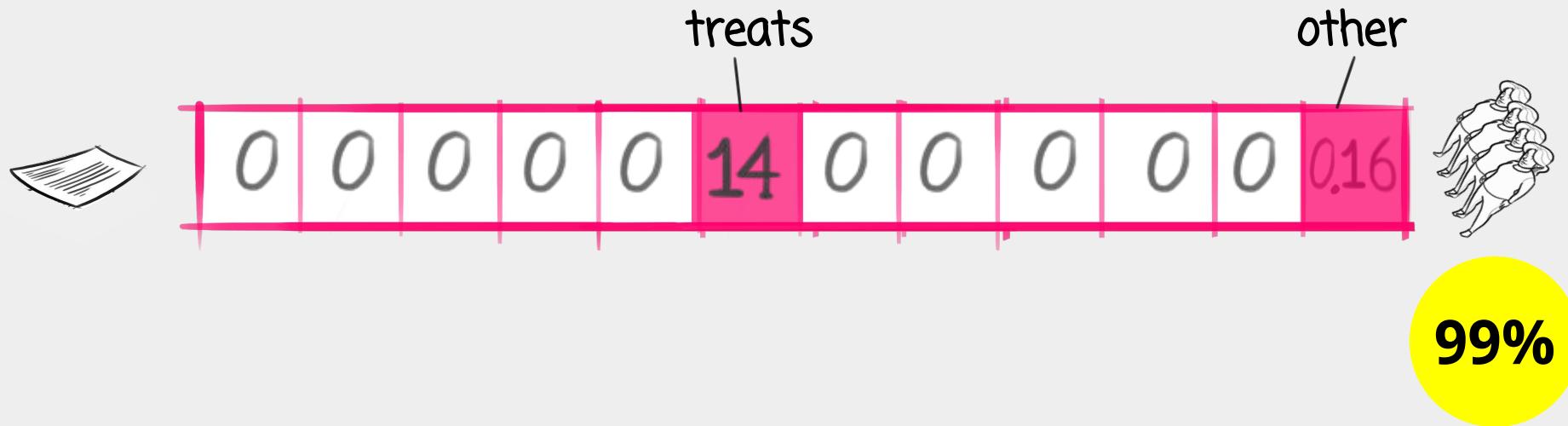
For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.



CROWDTRUTH METRICS

Applying all sides of the triangle

For prevention of malaria, use only in individuals traveling to
malarious areas where **CHLOROQUINE** resistant P.
falciparum **MALARIA** has not been reported.



CROWDTRUTH METRICS

Applying all sides of the triangle

APPLYING ALL SIDES OF THE TRIANGLE

Sentence quality:

overall worker agreement over one sentence

Worker quality:

overall agreement between one worker and the other workers

Annotation quality:

overall worker agreement over one annotation

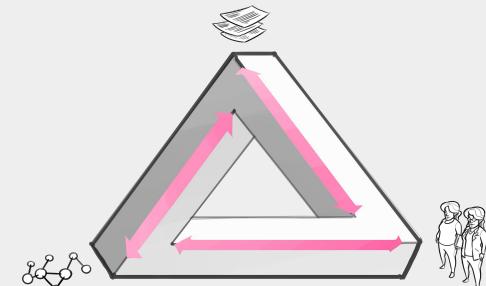
Calculated using **average cosine similarity** over **sentence vectors**.

Quality metrics are **inter-dependent**:

e.g. ambiguous sentences count less when determining worker quality, because disagreement is expected

CROWDTRUTH QUALITY METRICS

Ambiguity at any corner disseminates in the other corners



Frame-Sentence Score (FSS): the degree with which a particular frame matches the sense of the word in the sentence

$$FSS(s, f) = \frac{\sum \text{workers that picked } f \text{ in } s \cdot \text{worker quality}}{\sum \text{workers for } s \cdot \text{worker quality}}$$

Sentence Quality Score (SQS): overall worker agreement over one sentence, measured with cosine similarity

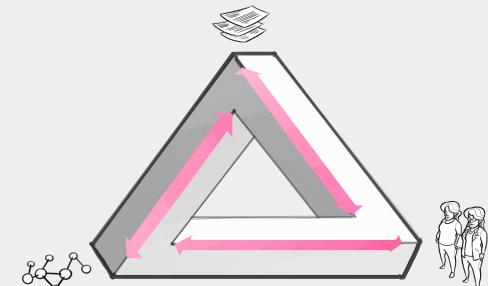
$$SQS(s) = \frac{\sum_{w1, w2 \in \text{workers for } s} \text{weighted cos sim}(\vec{w1}, \vec{w2}, FQS) \cdot \text{worker quality}(w1, w2)}{\sum_{w1, w2 \in \text{workers for } s} \text{worker quality}(w1, w2)}$$

Frame Quality Score (FQS): agreement over a frame in all sentences where the frame was picked at least once

$$FQS(f) = \frac{\sum_{FSS(s, f) > 0} FSS(s, f) SQS(s)}{\sum_{FSS(s, f) > 0} SQS(s)}$$

CROWDTRUTH QUALITY METRICS

Ambiguity at any corner disseminates in the other corners



1

Een Waisda? project

Uw bijdrage: 1

Man bijt hond Woordentikkertje



Selecteer een fragment om het spel mee te starten



Word geen slaaf van jezelf
6 keer gespeeld
hoogste score: 220



De goochelaar
0 keer gespeeld
hoogste score: 0



Jeugd indoor atletiek wedstrijd
0 keer gespeeld
hoogste score: 0

Topscorers deze week		
1	 marissa	20.745
2	 Willemien	9.595
3	 Ellen90	7.705
4	 lottebelice	5.155
5	 Julia N	4.840

Populairste tags deze week

man kinderen auto drink
 ncrv babbelbox hond
 geld lachen limburg
 steenoven groningen manbijt
 werk vrouw bijt verhalen
 gaarkeuken

Home

Een project van  en 

Mogelijk gemaakt door 

Select all tags that are relevant to the video

1 Watch the full video at least once

• Additional instructions here, only if strictly necessary



2 Select all tags that are explicitly shown in the video

Dover Alice Roberts Hovercraft France Cross-channel
 Journeys Calais Ocean Cruise Sea

3 Add new tags to the list:

my-tag

ADD +

4 Have you selected all relevant tags? Submit your result:

SUBMIT 2 TAGS >

VIDEO METADATA ENRICHMENT

The Netherlands Institute for Sound and Vision

On the role of user-generated metadata in audio visual collections (2011).

R. Gligorov, M. Hildebrand, J. van Ossenbruggen, G. Schreiber, L. Aroyo K-CAP2011

2

DIVE+

EXPLORATION PATH

- watersnoodramp
- 'Operatie Snel' ten bate van het Nationa...
- Watersnoodramp 1953
- De spoorwegverbinding door de overstr...
- Kruiningen
- Laatste stroomgat van Zuid-Beveland ...
- Herstel in van het ramgebi...
- Eerste grote succes bij de herstelwerkz...
- dijken 4
- overstroming 3

Laatste stroomgat van Zuid-Beveland gesloten ★

TITLE
Laatste stroomgat van Zuid-Beveland gesloten

ID
oai:openimages.eu:652021

DESCRIPTION
Bioscoopjournaals waarin Nederlandse onderwerpen van een bepaalde week worden gepresenteerd. Laatste grote reparatie van de door de Watersnoodramp aangerichte schade: een gat in een waterkering bij Kruiningen wordt gedicht. Een groot betonnen caisson wordt er bij dood tijdens een aantal sleepboten ingevoerd en daarna vastgezet. Archiefmateriaal: aftermath Watersnoodramp in deze omgeving.

DATE START
1953-07-01

DATE END

COPYRIGHT
<https://creativecommons.org/licenses/by-sa/3.0/nl/>

SOURCES
https://www.openbeelden.nl/feeds/oi/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:openimages.eu:652021

Related entities 7 / 10634

FILTER
example - exclude

DATE RANGE
No dates found

COLLECTION
 Unknown

	caissons		Watersnoodramp 1953		Kruiningen
	Nederland		Polygon-Profiliti (producent) / Nederlands Instituut voor Beeld en Geluid		creator of
	subject of		location for		published by
	6		2		3331
	subject of		location for		3906
	14		2		
	location for		3906		
	3331				

<https://www.openbeelden.nl/users/>

DIVE+

Explorative Search

DIVE into the event-based browsing of linked historical media (2015)

V De Boer, J Oomen, O Inel, L Aroyo, E Van Staveren, in Journal of Web Semantics:

3



With
Watson™

A close-up portrait of an elderly man with long, curly, light-colored hair. He has a serious expression and is looking slightly to the right. The background is dark and indistinct.

an artist or particular visitors

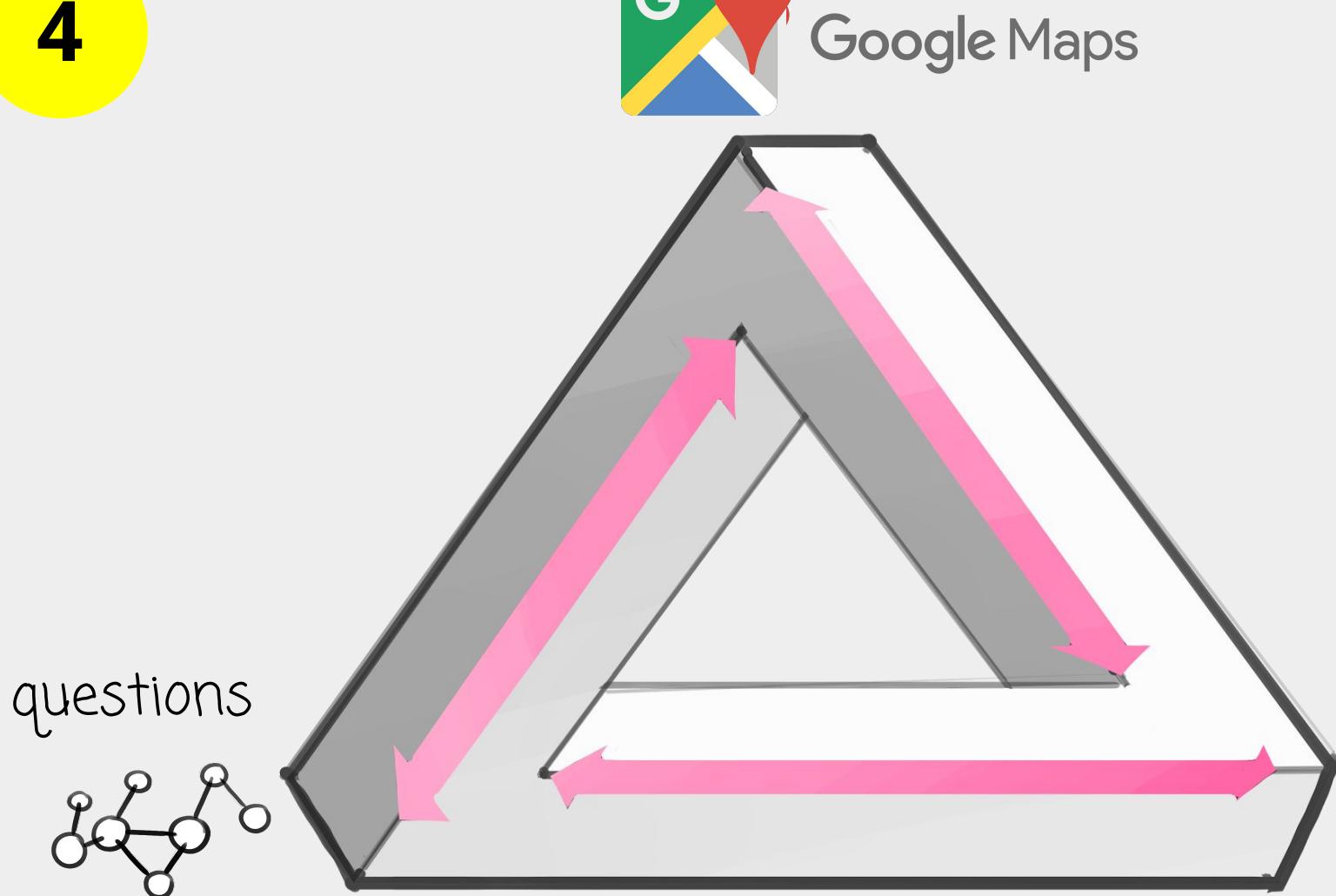
DEEP QA IN CULTURAL HERITAGE

Mauritshuis use case

4



Google Maps



CROWDTRUTH IN DEPLOYMENT

Google Maps

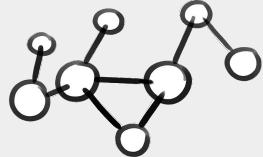
Google
Maps
reviewers

5



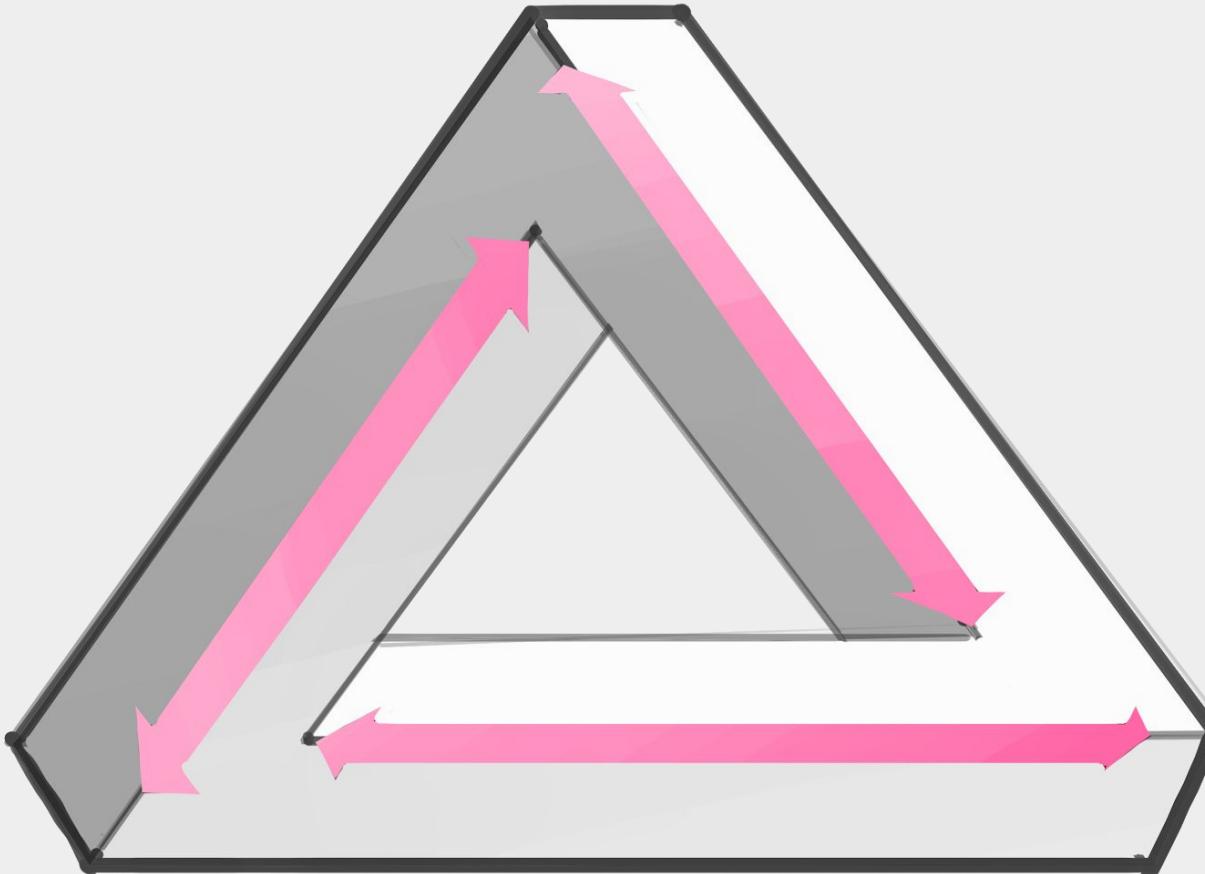
The
New York
Times

emotions



CROWDTRUTH IN DEPLOYMENT

New York Times



mTURK
crowd



RIJKS MUSEUM

COMMIT / A public- private
research community

TU Delft
Delft University of
Technology

VU UNIVERSITY
AMSTERDAM

CWI

SEALINC
Media

NICHESOURCING: FINDING NICHES IN THE CROWD

Accurator tool: SealincMedia Project

Chris Dijkshoorn, Victor De Boer, Lora Aroyo, Guus Schreiber (2014).
Accurator: Nichesourcing for Cultural Heritage

Select a recommended print to start

Landschap met een rivier

Printmaker: Nieulandt, Willem van (II)

Una et Eadem

Printmaker: Lairesse, Gerard de

Hanger met ineengesloten handen

Printmaker: anoniem

De vogels in 't woudt / Die zingen zeer stout: Om dat zy

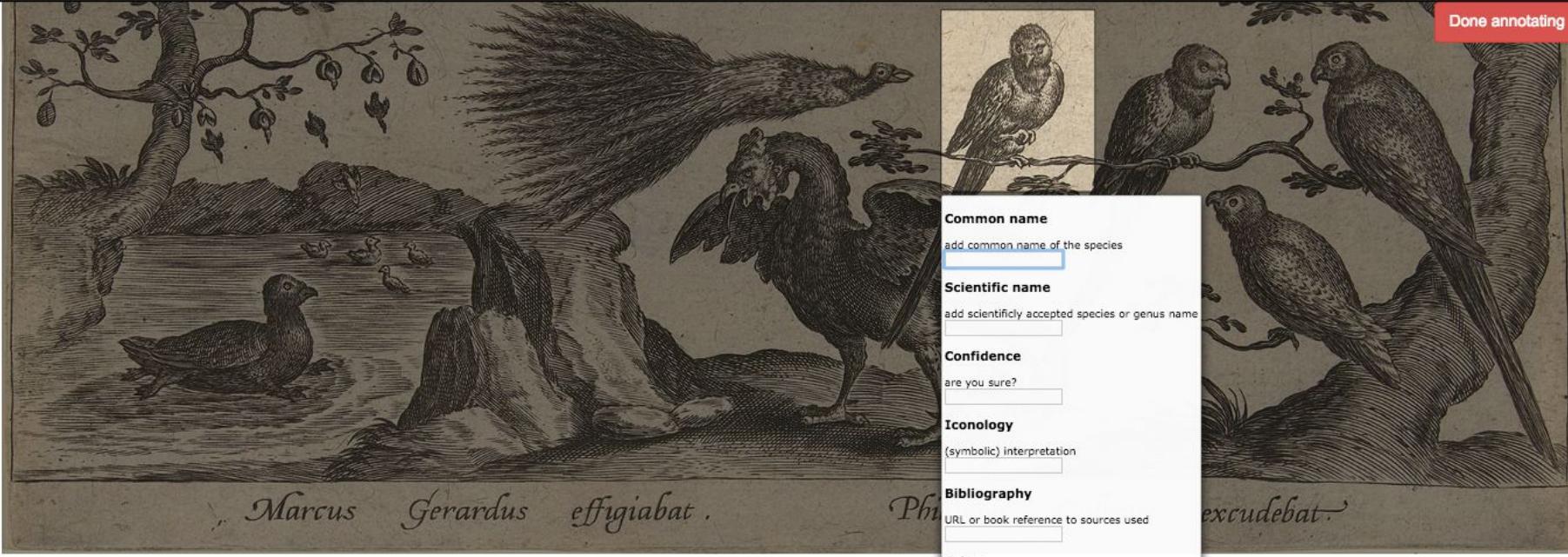
zingen met playzier / Daarom staan de vogels hier

Printmaker: anoniem

ACCURATOR

Menu

Done annotating



NICHESOURCING IN THE CULTURAL HERITAGE

Accurator tool

Chris Dijkshoorn, Victor De Boer, Lora Aroyo, Guus Schreiber (2014).
Accurator: Nichesourcing for Cultural Heritage

Select a recommended print to start

Landschap met een rivier

Printmaker: Nieulandt, Willem van (II)

Una et Eadem

Printmaker: Lairesse, Gerard de

Hanger met ineengesloten handen

Printmaker: anoniem

De vogels in 't woudt / Die zingen zeer stout: Om dat zy

zingen met playzier / Daarom staan de vogels hier

Printmaker: anoniem

ACCURATOR

Menu

Done annotating

Menu

ACCURATOR



Theme add the Biblical theme

Biblical person add the biblical person

Emotion which emotion is shown

Emotion intensity what is the intensity of the emotion?

Appraisal what appraisal can you discern?

References text/url

Done

Cancel

NICHE SOURCE

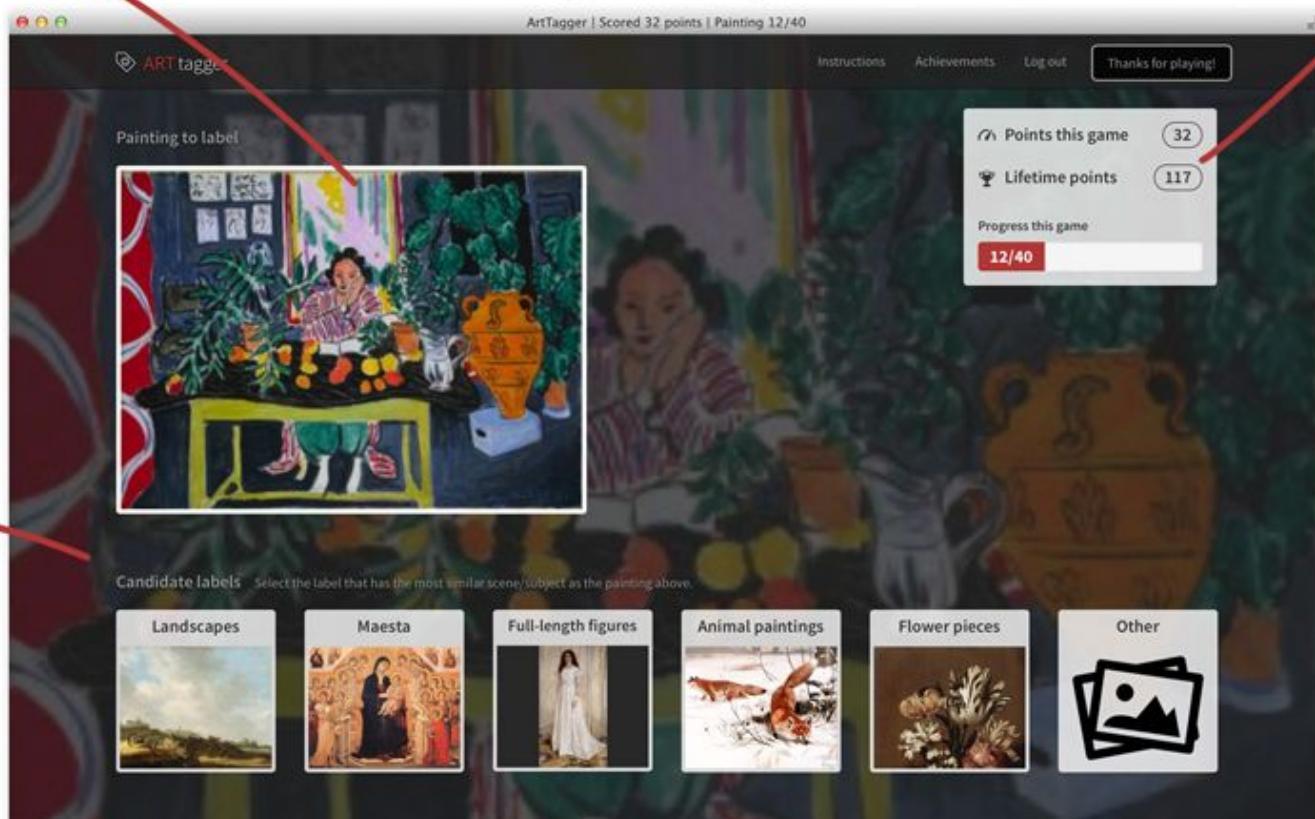
Accurator tool

Chris Dijkshoorn, Victor De Boer, Lora Aroyo, Guus Schreiber (2014).
Accurator: Nichesourcing for Cultural Heritage

Painting
to label

GAME PAGE

Points
to score



Candidate
labels

CREATING EXPERTS WITH GAMES

Accurator tool

Chris Dijkshoorn, Victor De Boer, Lora Aroyo, Guus Schreiber (2014).
Accurator: Nichesourcing for Cultural Heritage



Birdwatching Rijksmuseum

On Sunday October 4th, World Animal Day, the Rijksmuseum will host a birdwatching day in collaboration with Naturalis Biodiversity Center, Wikimedia Netherlands and the COMMIT SealincMedia project.

Rijksmuseum – Naturalis Biodiversity Center –
Wikimedia Nederland & COMMIT/ SEALINCMedia
present a unique birdwatching event

NICHESOURCING EVENTS

DigiBird Project

DigiBird: on the fly collection integration supported by the crowd (2017)

Chris Dijkshoorn, Christina-Lulia Bucur, Maarten Brinkerink, Sander Pieterse and Lora Aroyo

SOUND AND VISION



Naturalis
Biodiversity
Center



NICHESOURCING EVENTS

DigiBird Project

DigiBird: on the fly collection integration supported by the crowd (2017)

Chris Dijkshoorn, Christina-Lulia Bucur, Maarten Brinkerink, Sander Pieterse and Lora Aroyo

SOUND AND VISION





RIJKS STUDIO

Mogelijk gemaakt door **BankGiroLotterij**



Rembrandt Harmensz. van Rijn
RIJKS MUSEUM



Johannes Vermeer
RIJKS MUSEUM



Schilderijen
RIJKS MUSEUM



SUCCESS STORIES: RIJKSMUSEUM

Crowdsourcing with Riksstudio

632.953 artworks - 411.745 Riksstudios

FAIRview (Round 5)

6



NETHERLAND... PROJECT TYPE: PROTOTY... THE NETHERLANDS INSTITUTE FOR SOUND AND VISION...

Summary

The FAIRview project proposes an approach for adaptive news video summarization. Each video can have practically an infinite number of video summaries, each adapted to a specific set of user needs. For example, a given news video can generate a summary capturing all female anchors, or a summary capturing all the agricultural news items, or a summary capturing all the science news in this video, etc. Each of these video summaries will contain a portion of the original video content. In FAIRview, the selection of this portion is guided by a set of personalization features provided by users.

The solution

There's a growing demand from consumers to watch instead of reading news. To stay in tune with these changing viewing habits, news broadcasts are distributed in smaller units (atomised news). FAIRview presents a timely solution to this manual effort, by generating adaptive news video summaries suitable for distribution on social media, as well as on the archive's own video search environment. Secondly, as "fake videos" are on the rise as digital video manipulation techniques advance, FAIRview develops metrics to assess the quality of the video summaries and their potential for misinformation.

Digital News Innovation Fund

Google News Initiative

CROWDTRUTH IN DEPLOYMENT

Video Summarization

H2020 ReTV

Perspectives in Video Summarization

The image collage includes:

- A search results page for "Van Gogh movie" showing multiple video thumbnails.
- A diagram illustrating video summarization as a process of selecting key frames from a sequence.
- A network graph showing connections between video clips and characters from the TV show "Friends".
- A diagram showing a timeline of video clips with colored arrows indicating transitions or summarization steps.
- The ReTV logo with the tagline "Re-Inventing TV for the Interactive Age".
- A promotional slide for the ReTV project titled "From TV to ReTV: How television will Re-invent itself" with the URL <http://retv-project.eu/>.
- Social media icons for LinkedIn and Instagram, along with the project name "retv-project".
- The H2020 logo and the text "EU HORIZON 2020".
- A detailed diagram at the bottom right showing a timeline with specific "Relevant scenes" highlighted by red dashed lines and boxes.

Summarisation of multiple results is another way to cut the noise of the research into a more usable medium which allow us to visually explore a topic

Relevant scenes

Van Gogh movie

results

Van_gogh.mp4

Loving Vincent trailer

Vincent

15 Things You Didn't Know...

Vincent Van Gogh visit the...

Vincent Van Gogh - The Story

Summary of the result

Loving Vincent trailer

Van_gogh.mp4

Vincent Van Gogh - The Story

Vincent

IMDb

episodes

FRIENDS

Relevant scenes

<http://lora-aroyo.org> @laroyo



CaptureBias

- a theoretical framework for bias- and diversity-aware accuracy measures
- validate it in a range of domains, datasets
- demonstrate it with a proof of concept demo
- Gain understanding of:
 - the aspects of bias that find expression in video news
 - the aspects of bias that find expression in text news
 - the elements of storytelling that can be used for the transmission of bias
 - UI functionality that support social scientists and media study researchers to study bias
 - Providing larger context for individual items
 - Providing views over different perspectives
 - Stimulating pluralistic media consumption
 - how to combine bias in algorithms with bias in content presentation
- a set of open source 'diversity-aware' datasets, methods and algorithms

CAPTURING BIAS

Using Ambiguity to Support Media Scientists in News Videos Bias Detection



CaptureBias

Disagreement signals Ambiguity

Disagreement is caused by Ambiguity

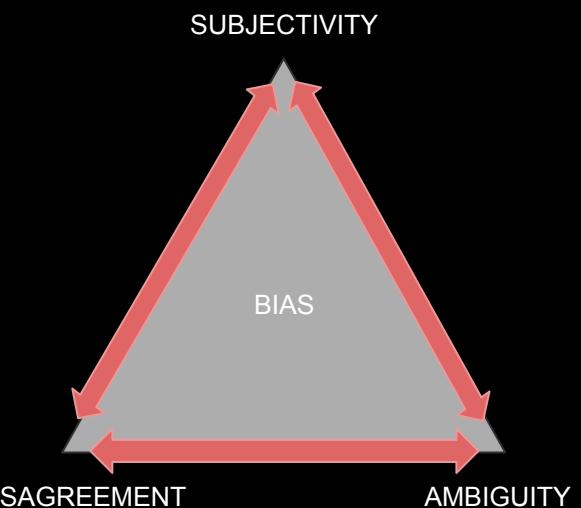
Framing will be manifested with high Ambiguity

Disagreement is caused by Subjectivity

Gender Bias interpretation depends on Subjectivity

Racial Bias interpretation depends on Subjectivity

Disagreement signal for Bias



CAPTURING BIAS

Using Ambiguity to Support Media Scientists in News Videos Bias Detection



LESSONS LEARNED ...

Crowds bring natural **diversity**

Crowds help gathering **real human semantics**

Crowds are **large** and contribute at **scale**

Crowds are **enthusiasts, motivated, driven by altruism**

There are **niches of experts** in the crowds

Experts and crowds are **complimentary**

Experts and crowds have **different semantics**

Experts and crowds use **different vocabularies**

Experts and crowds are **interested in different stories**

together they encompass a **multitude of opinions and perspectives**



The world is full
of **shades of grey**

Capturing and understanding opinions,
perspectives & contexts is in the center
of **understanding people**

CrowdTruth defines **hyper-dimensional space** to represent **ambiguity**

CrowdTruth defines **multi-dimensional space** to measure quality

Nichesourcing helps **expanding expertise**
beyond the walls of organizations

Nichesourcing needs **active engagement**
online and with onsite campaigns

LESSONS LEARNED ...

CROWDTRUTH.ORG

GITHUB.COM/CROWDTRUTH/CROWDTRUTH-CORE
PYPI.ORG/PROJECT/CROWDTRUTH

DATA.CROWDTRUTH.ORG

<http://lora-aroyo.org> @laroyo

