# What Do NLP Researchers Believe?
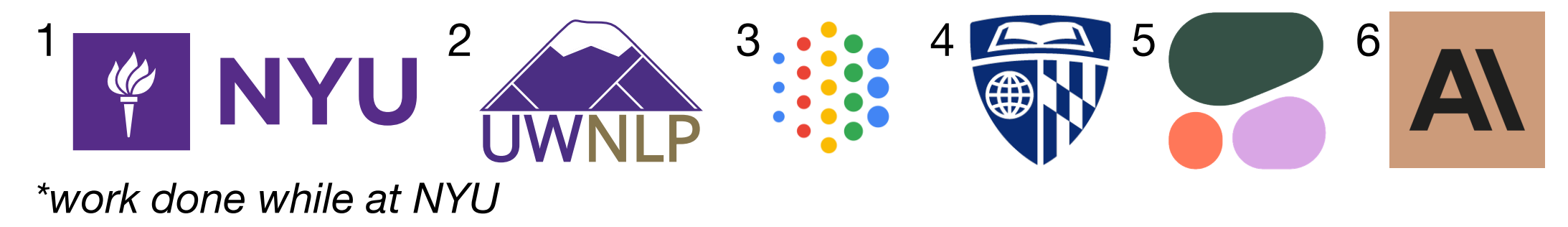## Results of the NLP Community Metasurvey

Julian Michael,[1] Ari Holtzman,[2] Alicia Parrish,[3]* Aaron Mueller,[4] Alex Wang,[5]*
Angelica Chen,[1] Divyam Madaan,[1] Nikita Nangia,[1]
Richard Yuanzhe Pang,[1] Jason Phang,[1]
Samuel R. Bowman[1,6]*
https://nlpsurvey.net/results

[1] NYU  [2] UWNLP  [3]  [4]  [5]  [6] AI
*work done while at NYU

### Motivation

- Developments in NLP are **hard to predict** & **becoming higher stakes.**
- This raises **a lot of questions** where there is **a lot of disagreement.**
- It's hard to know via social media what most people **actually think.**
- This makes it **harder to communicate and build shared agendas.**
  - Can lead researchers to avoid promising topics, waste effort, and argue on the basis of assumptions not held by their audience.
  - Worst case scenario: **a fractured discourse**
- We can do something about this — **run a survey!**

### Methodology

- Each question is an opinion with a 4-point Likert scale plus 3 options:

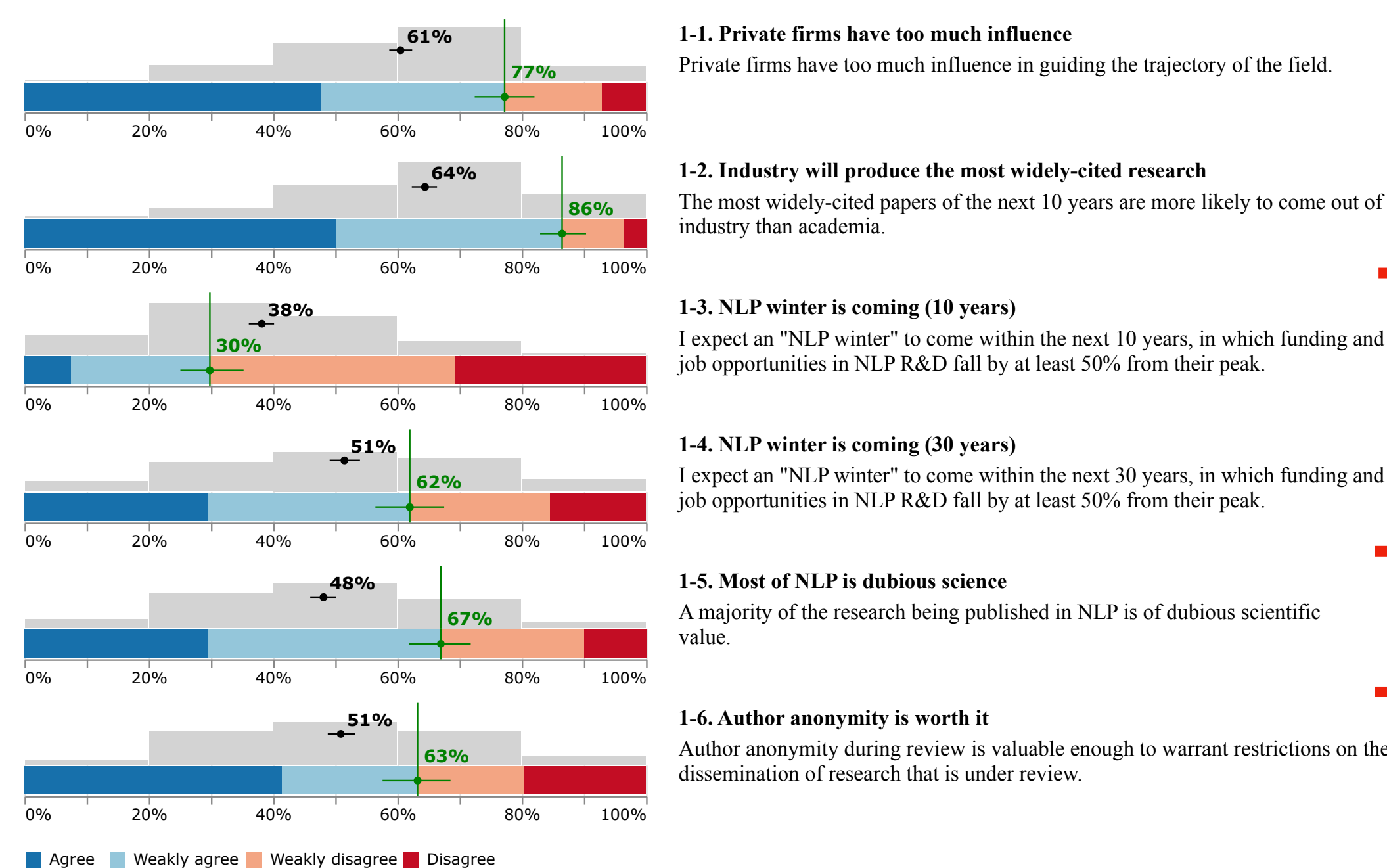> **Q1-1.** *Private firms have too much influence in guiding the trajectory of the field.*
>
> - Agree
> - Weakly agree
> - Weakly disagree
> - Disagree
>
> *Other:*
> - Insufficiently informed on the issue
> - Question is ill-posed
> - Prefer not to say

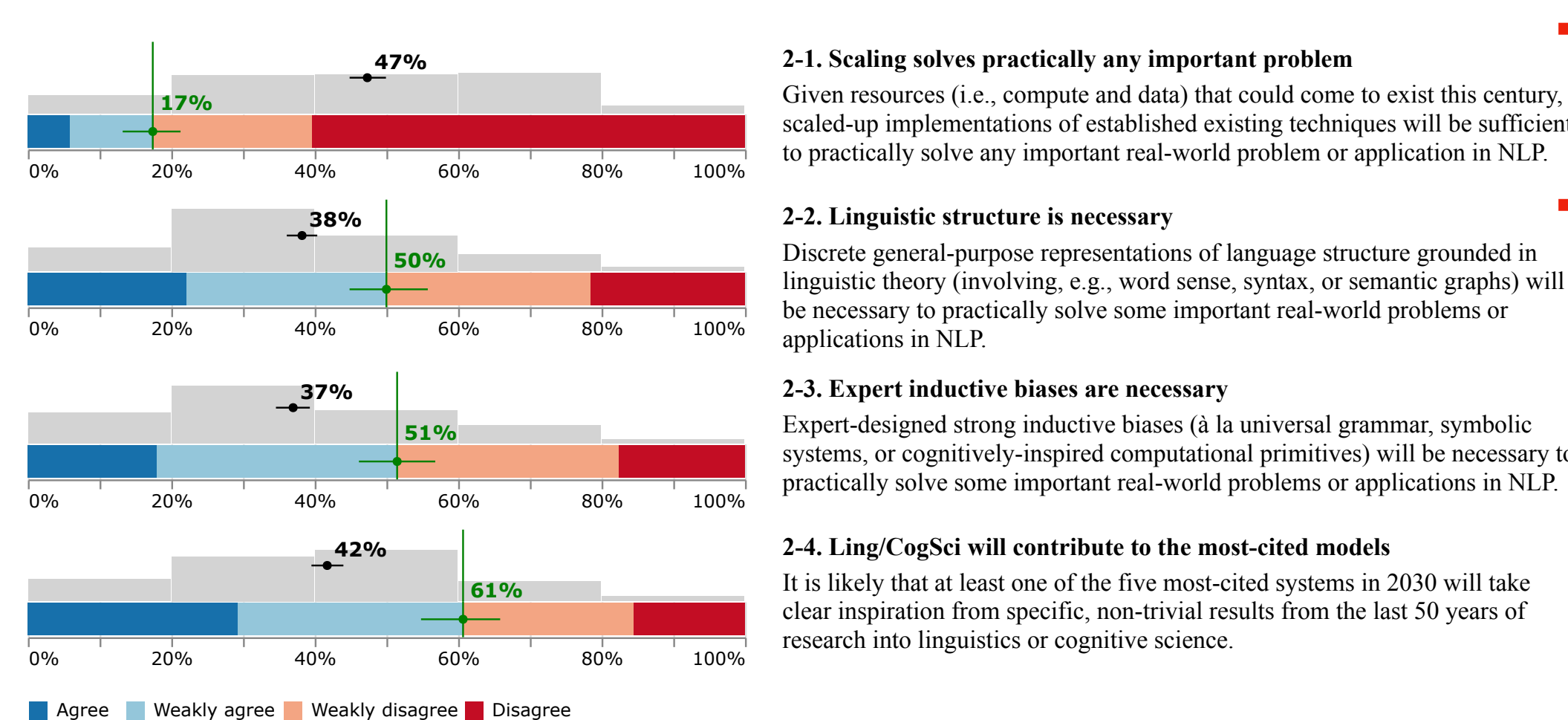- Then respondents *predict what others will say*, and we can compare!

> *Of those on the agree/disagree spectrum, what percentage of community members do you think will mark "agree" or "weakly agree"?*
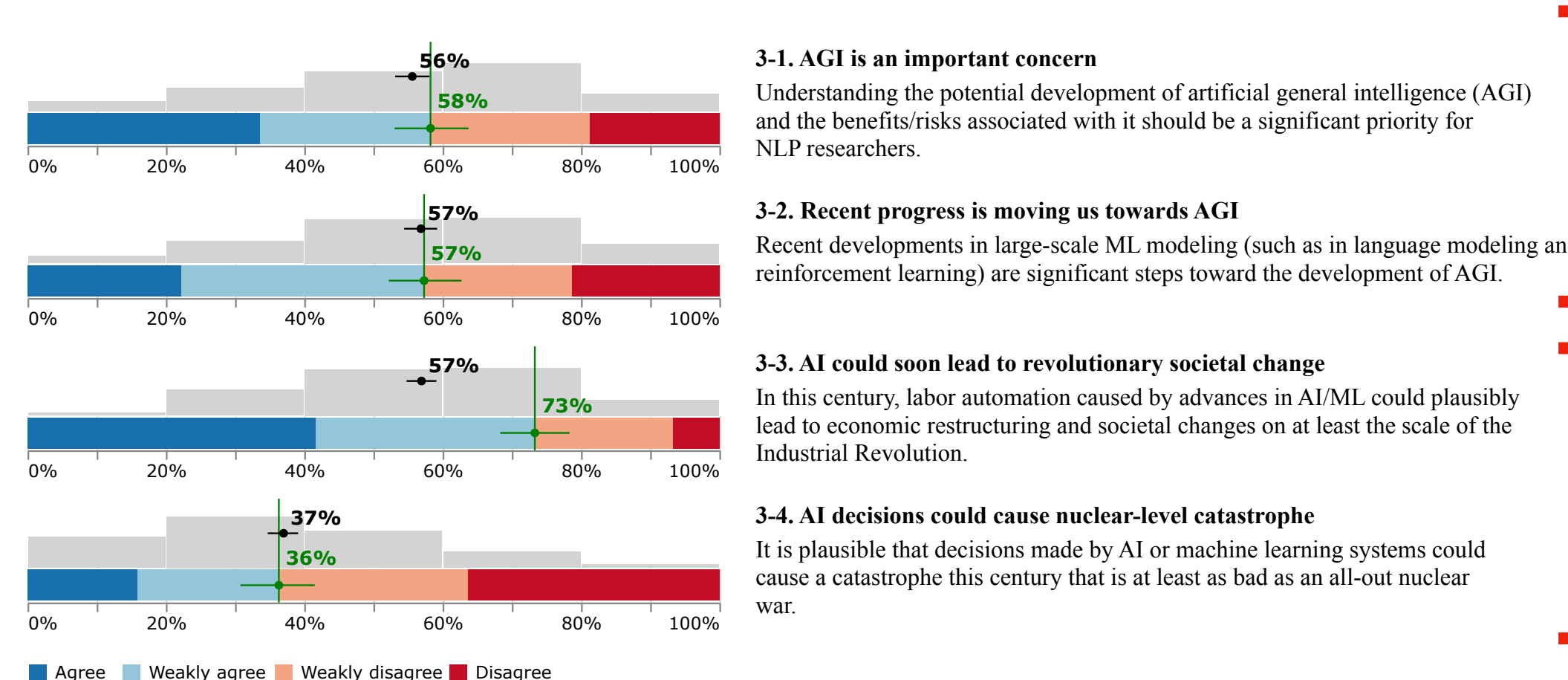>
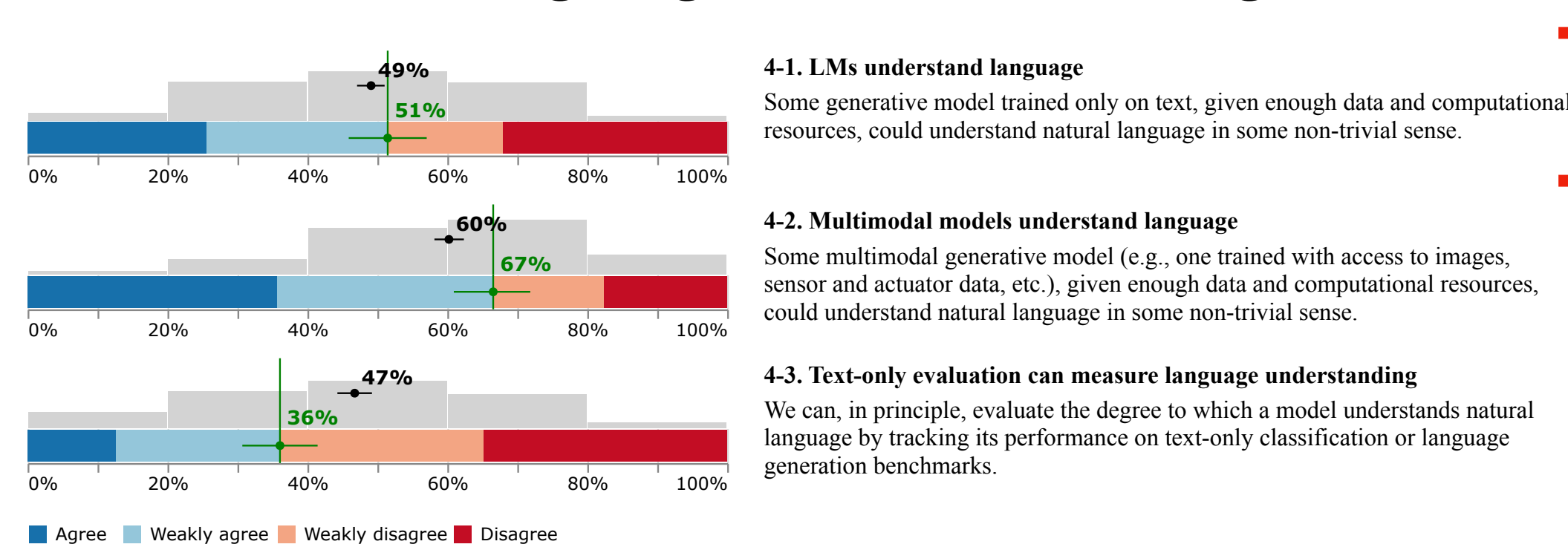> 0–20%  20–40%  40–60%  60–80%  80–100%

### Analyze the results yourself on our dashboard!

### Observations

### Main Takeaways

- Scaling maximalism was a surprisingly rare view in early/mid 2022
- Surprisingly many see a role for increased use of expert-designed inductive biases, linguistic theory, and interdisciplinary work
- There is little consensus on important ethical issues (e.g., misuse, environmental impact, regulation)
- A significant fraction (36%) see catastrophic risk from AI as plausible, and a majority (74%) see AI as plausibly transformative for society

### Demographics

- **327 respondents** in our target demographic (≥2 *CL pubs 2019–2022)



## 1. State of the Field



**1-1. Private firms have too much influence**
Private firms have too much influence in guiding the trajectory of the field.

**1-2. Industry will produce the most widely-cited research**
The most widely-cited papers of the next 10 years are more likely to come out of industry than academia.

**1-3. NLP winter is coming (10 years)**
I expect an "NLP winter" to come within the next 10 years, in which funding and job opportunities in NLP R&D fall by at least 50% from their peak.

**1-4. NLP winter is coming (30 years)**
I expect an "NLP winter" to come within the next 30 years, in which funding and job opportunities in NLP R&D fall by at least 50% from their peak.

**1-5. Most of NLP is dubious science**
A majority of the research being published in NLP is of dubious scientific value.

**1-6. Author anonymity is worth it**
Author anonymity during review is valuable enough to warrant restrictions on the dissemination of research that is under review.

## 2. Scale, Inductive Bias, and Adjacent Fields



**2-1. Scaling solves practically any important problem**
Given resources (i.e., compute and data) that could come to exist this century, scaled-up implementations of established existing techniques will be sufficient to practically solve any important real-world problem or application in NLP.

**2-2. Linguistic structure is necessary**
Discrete general-purpose representations of language structure grounded in linguistic theory (involving, e.g., word sense, syntax, or semantic graphs) will be necessary to practically solve some important real-world problems or applications in NLP.

**2-3. Expert inductive biases are necessary**
Expert-designed strong inductive biases (à la universal grammar, symbolic systems, or cognitively-inspired computational primitives) will be necessary to practically solve some important real-world problems or applications in NLP.

**2-4. Ling/CogSci will contribute to the most-cited models**
It is likely that at least one of the five most-cited systems in 2030 will take clear inspiration from specific, non-trivial results from the last 50 years of research into linguistics or cognitive science.

## 3. AGI and Major Risks



**3-1. AGI is an important concern**
Understanding the potential development of artificial general intelligence (AGI) and the benefits/risks associated with it should be a significant priority for NLP researchers.

**3-2. Recent progress is moving us towards AGI**
Recent developments in large-scale ML modeling (such as in language modeling and reinforcement learning) are significant steps toward the development of AGI.

**3-3. AI could soon lead to revolutionary societal change**
In this century, labor automation caused by advances in AI/ML could plausibly lead to economic restructuring and societal changes on at least the scale of the Industrial Revolution.

**3-4. AI decisions could cause nuclear-level catastrophe**
It is plausible that decisions made by AI or machine learning systems could cause a catastrophe this century that is at least as bad as an all-out nuclear war.

## 4. Language Understanding



**4-1. LMs understand language**
Some generative model trained only on text, given enough data and computational resources, could understand language in some non-trivial sense.

**4-2. Multimodal models understand language**
Some multimodal generative model (e.g., one trained with access to images, sensor and actuator data, etc.), given enough data and computational resources, could understand natural language in some non-trivial sense.

**4-3. Text-only evaluation can measure language understanding**
We can, in principle, evaluate the degree to which a model understands natural language by tracking its performance on text-only classification or language generation benchmarks.

### Observations

#### Respondents skew senior, academic, and online
A plurality (41%) of respondents are senior, 73% work in academia, and 89% publish for their job. Most (76%) use Twitter.

#### The USA is overrepresented
The USA is overrepresented (58%, compared to 35% of ACL members), and Asia underrepresented but the gender ratio tracks available ACL statistics.

#### Industry is seen as influential
Industry is seen as highly influential — 77% think it has too much influence, and 86% think it will produce the most widely-cited research. The prevalence of these views are underestimated by 15–20%.

#### NLP seen as dubious science
67% believe that a majority of NLP publications have dubious scientific value.

#### People want to prioritize interdisciplinary insights
82% believe we should incorporate more insights from adjacent fields / domain sciences — many more than the 53% predicted.

#### Scaling maximalism is rare
Only 17% believe in *scaling maximalism* — the view that scale will solve all practical problems. The view's prevalence is greatly overestimated, at 47%.

#### Problem formulation and task design: a frontier?
The most influential advances of the next 10 years are expected to be in *problem formulation and task design* rather than scale, hardware, data, or algorithms.

#### AGI and "understanding" are controversial still
Whether AGI is an important concern, and whether LMs understand language, are split down the middle.

#### AI is seen as plausibly having massive impact on society
73% think AI could plausibly bring revolutionary societal change, and 36% think catastrophic risk from AI is plausible, even though a vast majority (87–89%) thinks NLP is net good for the world.

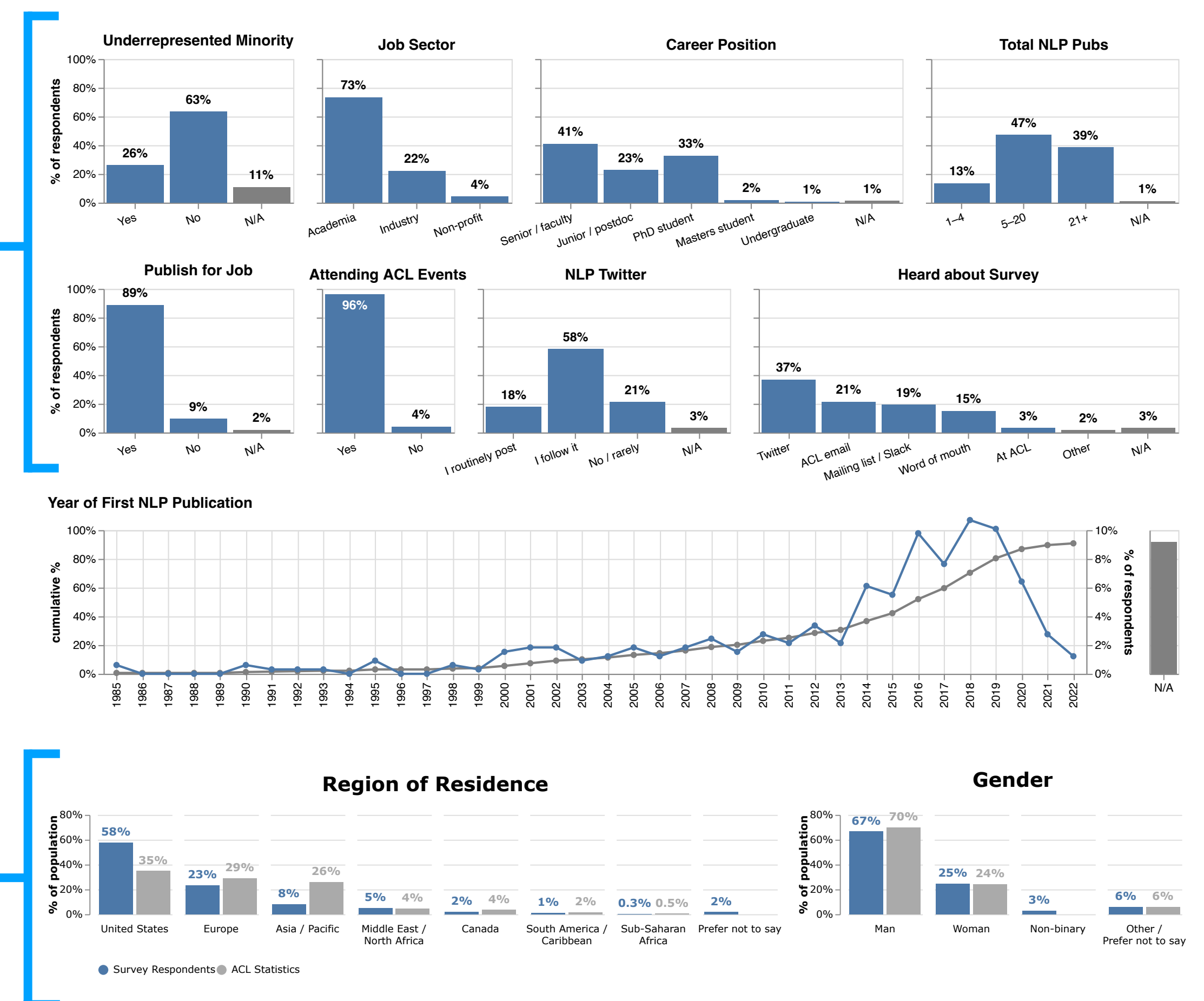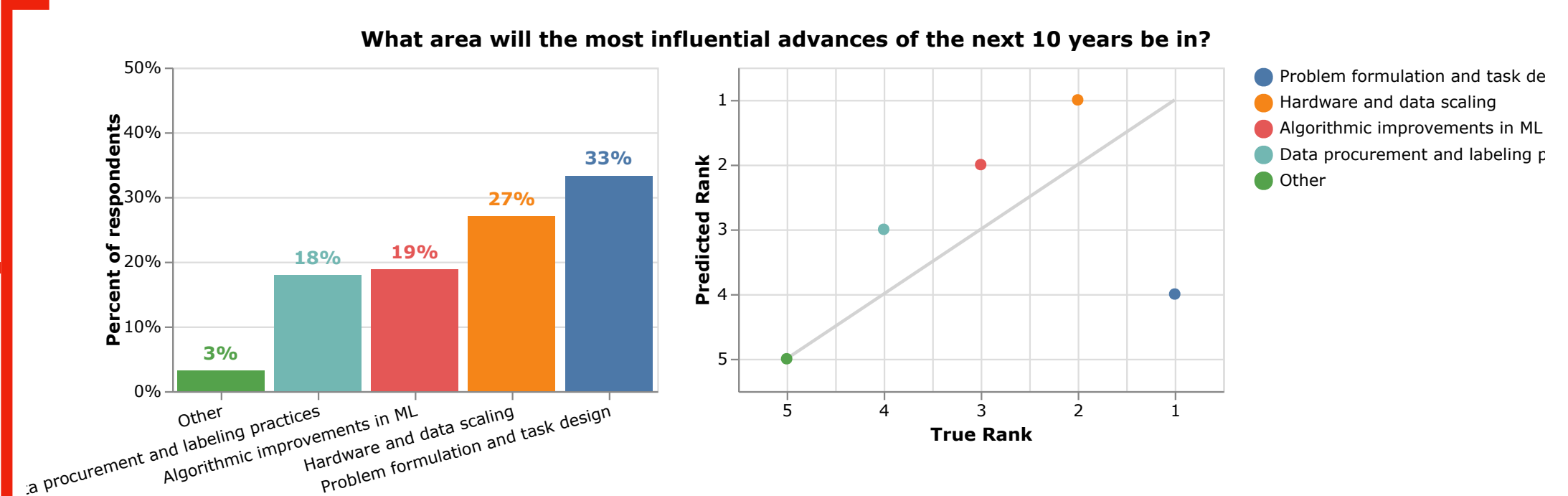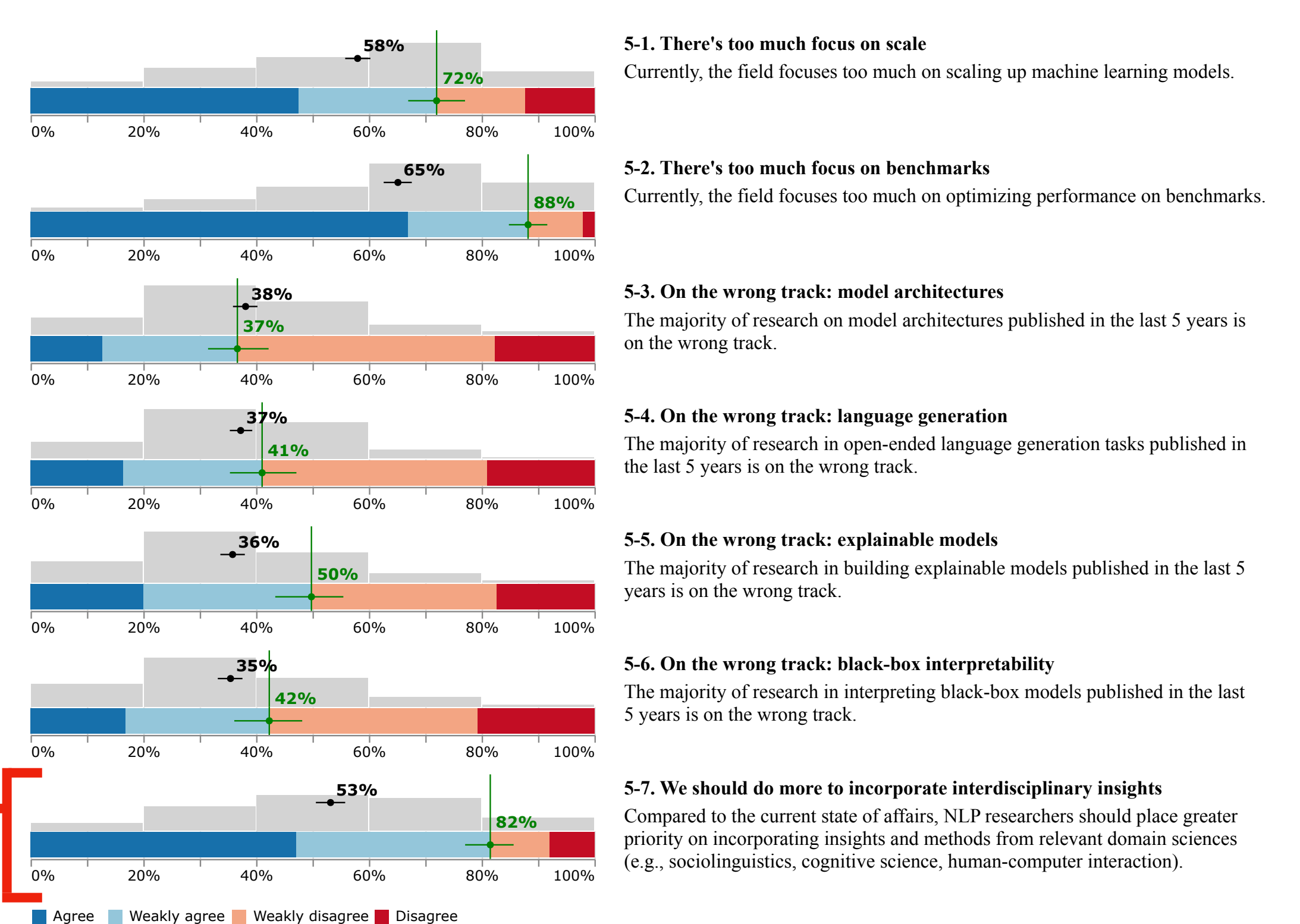#### There is concern for carbon but skepticism of regulation
60% of respondents think carbon footprint is a major concern, but less than half think NLP should be regulated by governments.

## 5. Promising Research Directions



**5-1. There's too much focus on scale**
Currently, the field focuses too much on scaling up machine learning models.

**5-2. There's too much focus on benchmarks**
Currently, the field focuses too much on optimizing performance on benchmarks.

**5-3. On the wrong track: model architectures**
The majority of research on model architectures published in the last 5 years is on the wrong track.

**5-4. On the wrong track: language generation**
The majority of research in open-ended language generation tasks published in the last 5 years is on the wrong track.

**5-5. On the wrong track: explainable models**
The majority of research in building explainable models published in the last 5 years is on the wrong track.

**5-6. On the wrong track: black-box interpretability**
The majority of research in interpreting black-box models published in the last 5 years is on the wrong track.

**5-7. We should do more to incorporate interdisciplinary insights**
Compared to the current state of affairs, NLP researchers should place greater priority on incorporating insights and methods from relevant domain sciences (e.g., sociolinguistics, cognitive science, human-computer interaction).



## 6. Ethics



**6-1. NLP's past net impact is good**
On net, NLP research has had a positive impact on the world.

**6-2. NLP's future net impact is good**
On net, NLP research continuing into the future will have a positive impact on the world.

**6-3. It is unethical to build easily-misusable systems**
It is unethical to build and publicly release a system which can easily be used in harmful ways.

**6-4. Ethical and scientific considerations can conflict**
In the context of NLP research, ethical considerations can sometimes be at odds with the progress of science.

**6-5. Ethical concerns mostly reduce to data quality and model accuracy**
The main ethical challenges posed by current ML systems can, in principle, be solved through improvements in data quality/coverage and model accuracy.

**6-6. It is unethical to predict psychological characteristics**
It is inherently unethical to develop ML systems for predicting people's internal psychological characteristics (e.g., emotions, gender identity, sexual orientation).

**6-7. Carbon footprint is a major concern**
The carbon footprint of training large models should be a major concern for NLP researchers.

**6-8. NLP should be regulated**
The development and deployment of NLP systems should be regulated by governments.