

# CHAPTER 6: Diversity in AI



Artificial Intelligence  
Index Report 2021



## CHAPTER 6: Chapter Preview

Overview	3	New Computing PhDs in the United States by Race/Ethnicity	11
Chapter Highlights	4	CS Tenure-Track Faculty by Race/Ethnicity	12
<b>6.1 GENDER DIVERSITY IN AI</b>	<b>5</b>	Black in AI	12
Women in Academic AI Settings	5	<b>6.3 GENDER IDENTITY AND SEXUAL ORIENTATION IN AI</b>	<b>13</b>
Women in the AI Workforce	6	Queer in AI	13
Women in Machine Learning Workshops	7	Demographics Breakdown	13
Workshop Participants	7	Experience as Queer Practitioners	15
Demographics Breakdown	8	<b>APPENDIX</b>	<b>17</b>
<b>6.2 RACIAL AND ETHNIC DIVERSITY IN AI</b>	<b>10</b>		
New AI PhDs in the United States by Race/Ethnicity	10		

**ACCESS THE PUBLIC DATA**



# Overview

While artificial intelligence (AI) systems have the potential to dramatically affect society, the people building AI systems are not representative of the people those systems are meant to serve. The AI workforce remains predominantly male and lacking in diversity in both academia and the industry, despite many years highlighting the disadvantages and risks this engenders. The lack of diversity in race and ethnicity, gender identity, and sexual orientation not only risks creating an uneven distribution of power in the workforce, but also, equally important, reinforces existing inequalities generated by AI systems, reduces the scope of individuals and organizations for whom these systems work, and contributes to unjust outcomes.

This chapter presents diversity statistics within the AI workforce and academia. It draws on collaborations with various organizations—in particular, Women in Machine Learning (WiML), Black in AI (BAI), and Queer in AI (QAI)—each of which aims to improve diversity in some dimension in the field. The data is neither comprehensive nor conclusive. In preparing this chapter, the AI Index team encountered significant challenges as a result of the sparsity of publicly available demographic data. The lack of publicly available demographic data limits the degree to which statistical analyses can assess the impact of the lack of diversity in the AI workforce on society as well as broader technology development. The diversity issue in AI is well known, and making more data available from both academia and industry is essential to measuring the scale of the problem and addressing it.

There are many dimensions of diversity that this chapter does not cover, including AI professionals with disabilities; nor does it consider diversity through an intersectional lens. Other dimensions will be addressed in future iterations of this report. Moreover, these diversity statistics tell only part of the story. The daily challenges of minorities and marginalized groups working in AI, as well as the structural problems within organizations that contribute to the lack of diversity, require more extensive data collection and analysis.

---

<sup>1</sup> We thank Women in Machine Learning, Black in AI, and Queer in AI for their work to increase diversity in AI, for sharing their data, and for partnering with us.

## CHAPTER HIGHLIGHTS

- The percentages of female AI PhD graduates and tenure-track computer science (CS) faculty have remained low for more than a decade. Female graduates of AI PhD programs in North America have accounted for less than 18% of all PhD graduates on average, according to an annual survey from the Computing Research Association (CRA). An AI Index survey suggests that female faculty make up just 16% of all tenure-track CS faculty at several universities around the world.
- The CRA survey suggests that in 2019, among new U.S. resident AI PhD graduates, 45% were white, while 22.4% were Asian, 3.2% were Hispanic, and 2.4% were African American.
- The percentage of white (non-Hispanic) new computing PhDs has changed little over the last 10 years, accounting for 62.7% on average. The share of Black or African American (non-Hispanic) and Hispanic computing PhDs in the same period is significantly lower, with an average of 3.1% and 3.3%, respectively.
- The participation in Black in AI workshops, which are co-located with the Conference on Neural Information Processing Systems (NeurIPS), has grown significantly in recent years. The numbers of attendees and submitted papers in 2019 are 2.6 times higher than in 2017, while the number of accepted papers is 2.1 times higher.
- In a membership survey by Queer in AI in 2020, almost half the respondents said they view the lack of inclusiveness in the field as an obstacle they have faced in becoming a queer practitioner in the AI/ML field. More than 40% of members surveyed said they have experienced discrimination or harassment as a queer person at work or school.

# 6.1 GENDER DIVERSITY IN AI

## WOMEN IN ACADEMIC AI SETTINGS

Chapter 4 introduced the AI Index survey that evaluates the state of AI education in CS departments at top universities around the world, along with the Computer Research Association’s annual Taulbee Survey on the enrollment, production, and employment of PhDs in information, computer science, and computer engineering in North America.

Data from both surveys show that the percentage of female AI and CS PhD graduates as well as tenure-track CS faculty remains low. Female graduates of AI PhD programs and CS PhD programs have accounted for 18.3% of all PhD graduates on average within the past 10 years (Figure 6.1.1). Among the 17 universities that completed the AI Index survey of CS programs globally, female faculty make up just 16.1% of all tenure-track faculty whose primary research focus area is AI (Figure 6.1.2).

TENURE-TRACK FACULTY at CS DEPARTMENTS of TOP UNIVERSITIES around the WORLD by GENDER, AY 2019-20

Source: AI Index, 2020 | Chart: 2021 AI Index Report

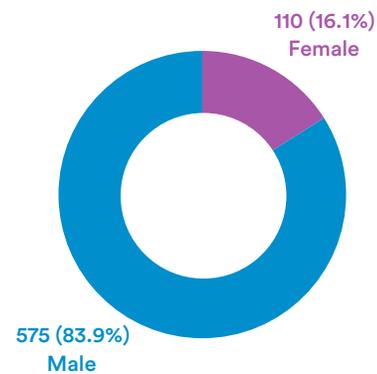


Figure 6.1.2

FEMALE NEW AI and CS PHDS (% of TOTAL NEW AI and CS PHDS) in NORTH AMERICA, 2010-19

Source: CRA Taulbee Survey, 2020 | Chart: 2021 AI Index Report

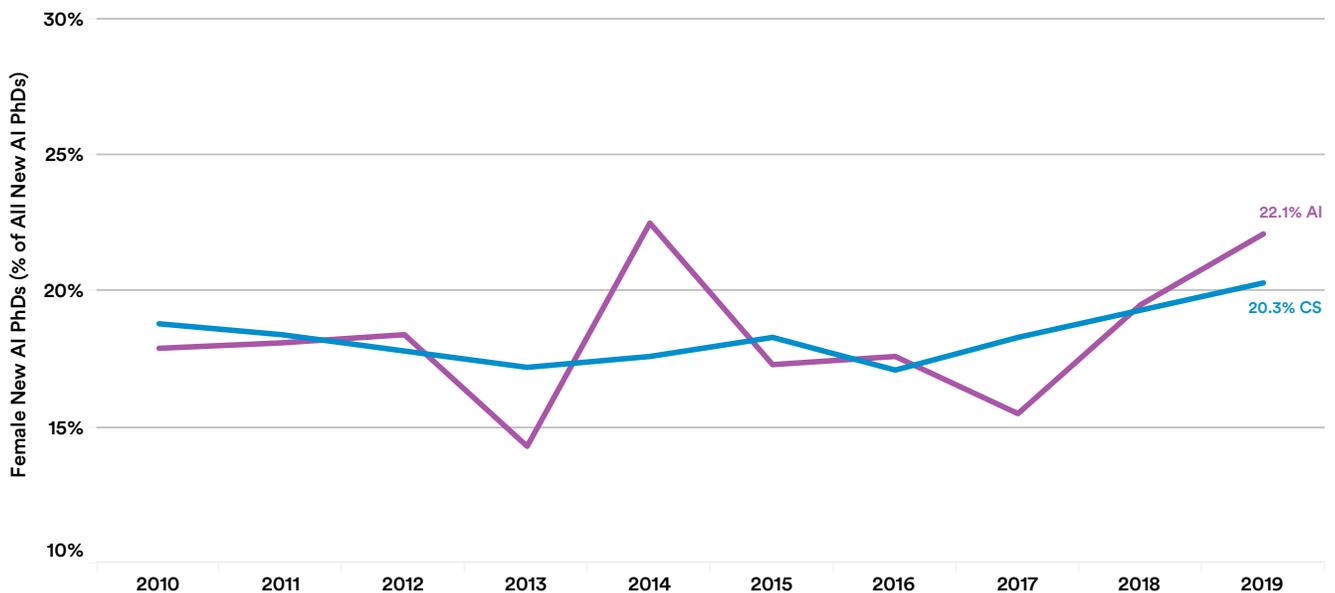


Figure 6.1.1

## WOMEN IN THE AI WORKFORCE

Chapter 3 introduced the “global relative AI skills penetration rate,” a measure that reflects the prevalence of AI skills across occupations, or the intensity with which people in certain occupations use AI skills. Figure 6.1.3 shows AI skills penetration by country for female and male labor pools in a set of select countries.<sup>2</sup> The data suggest that across the majority of these countries, the AI skills penetration rate for women is lower than that for men. Among the 12 countries we examined, India, South Korea, Singapore, and Australia are the closest to reaching equity in terms of the AI skills penetration rate of females and males.

**This data suggests that across the majority of select countries, the AI skills penetration rate for women is lower than it is for men.**

### RELATIVE AI SKILLS PENETRATION RATE by GENDER, 2015-20

Source: LinkedIn, 2020 | Chart: 2021 AI Index Report

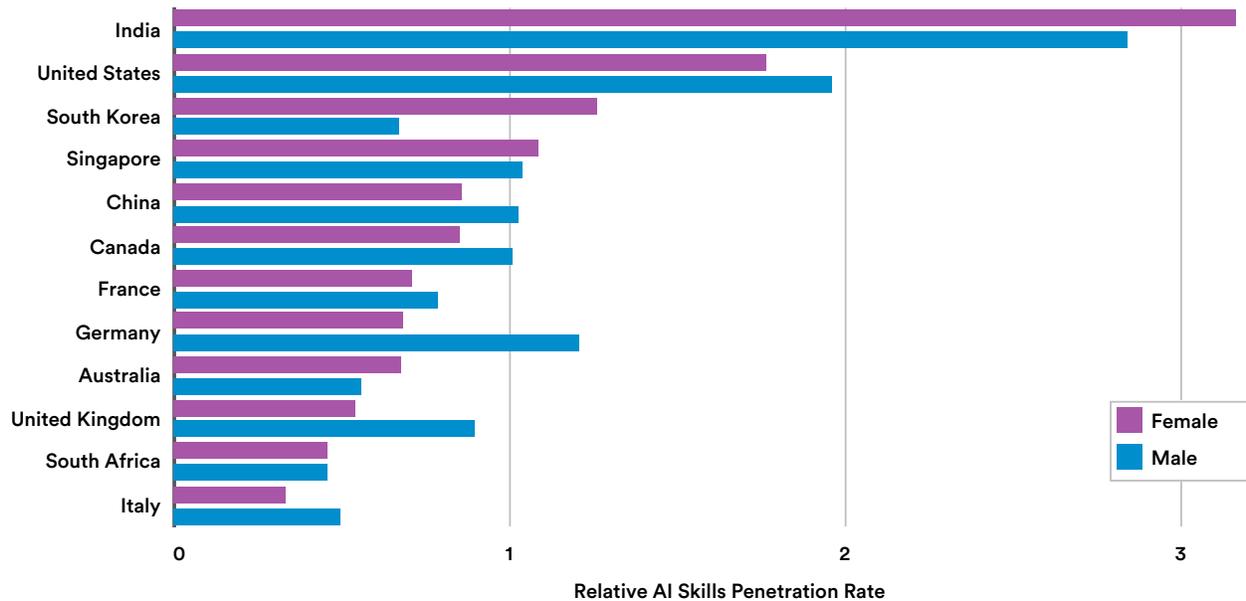


Figure 6.1.3

<sup>2</sup> Countries included are a select sample of eligible countries with at least 40% labor force coverage by LinkedIn and at least 10 AI hires in any given month. China and India were included in this sample because of their increasing importance in the global economy, but LinkedIn coverage in these countries does not reach 40% of the workforce. Insights for these countries may not provide as full a picture as other countries, and should be interpreted accordingly.

## WOMEN IN MACHINE LEARNING WORKSHOPS

Women in Machine Learning, founded in 2006 by Hanna Wallach, Jenn Wortman, and Lisa Wainer, is an organization that runs events and programs to support women in the field of machine learning (ML). This section presents statistics from its annual technical workshops, which are held at NeurIPS. In 2020, WiML also hosted for the first time a full-day “Un-Workshop” at the International Conference on Machine Learning 2020, which drew 812 participants.

### Workshop Participants

The number of participants attending WiML workshops at NeurIPS has been steadily increasing since the workshops were first offered in 2006. According to the organization, the WiML workshop in 2020 was completely virtual

because of the pandemic and delivered on a new platform (Gather.Town); these two factors may make attendance numbers harder to compare to those of previous years. Figure 6.1.4 shows an estimate of 925 attendees in 2020, based on the number of individuals who accessed the virtual platform.

In the past 10 years, WiML workshops have expanded their programs to include mentoring roundtables, where more senior participants offer one-on-one feedback and professional advice, in addition to the main session that includes keynotes and poster presentations. Similar opportunities may have contributed to the increase in attendance since 2014. Between 2016 and 2019, the WiML workshop attendance is on average about 10% of the overall NeurIPS attendance.

### NUMBER of PARTICIPANTS at WIML WORKSHOP at NEURIPS, 2006-20

Source: Women in Machine Learning, 2020 | Chart: 2021 AI Index Report

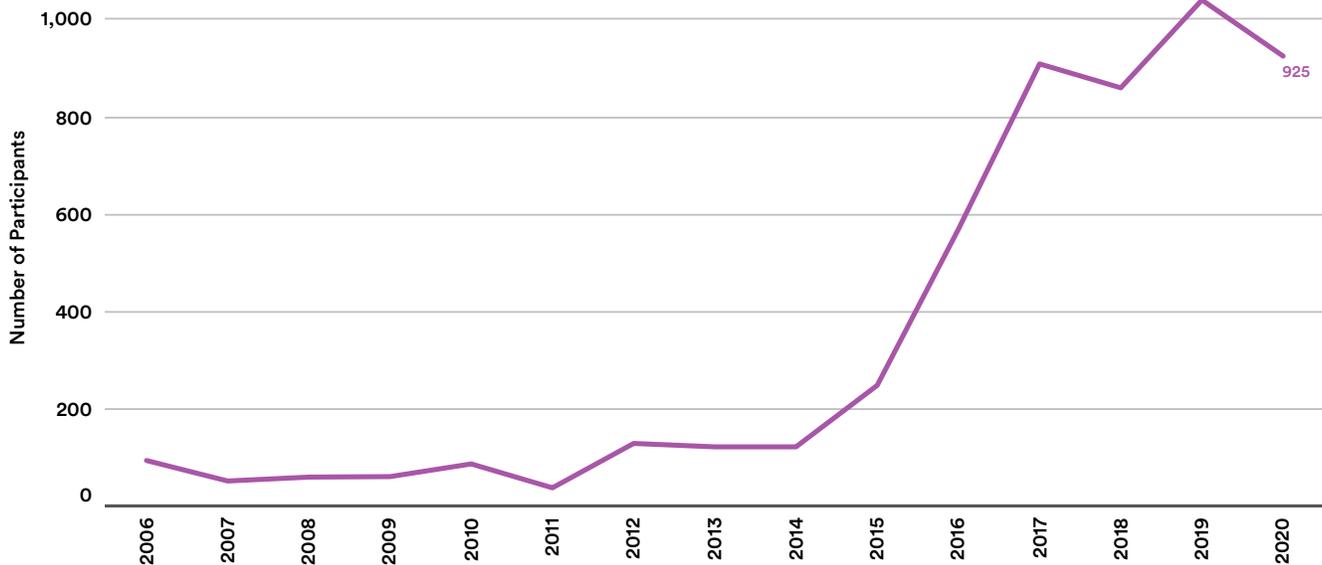


Figure 6.1.4

## Demographics Breakdown

The following geographic, professional position, and gender breakdowns are based only on participants at the 2020 WiML workshop at NeurIPS who consented to having the information aggregated and who spent at least 10 minutes on the virtual platform through which the workshop was offered. Among the participants, 89.5% were women and/or nonbinary, 10.4% were men (Figure 6.1.5), and a large majority were from North America (Figure 6.1.6). Further, as shown in Figure 6.1.7, students—including PhD, master’s, and undergraduate students—make up more than half the participants (54.6%). Among participants who work in the industry, research scientist/engineer and data scientist/engineer are the most commonly held professional positions.

**Among the participants, 89.5% were women and/or nonbinary, 10.4% were men, and a large majority were from North America. Further, students—including PhD, master’s, and undergraduate students—make up more than half the participants (54.6%).**

### PARTICIPANTS of WiML WORKSHOP at NEURIPS (% of TOTAL) by GENDER, 2020

Source: Women in Machine Learning, 2020 | Chart: 2021 AI Index Report

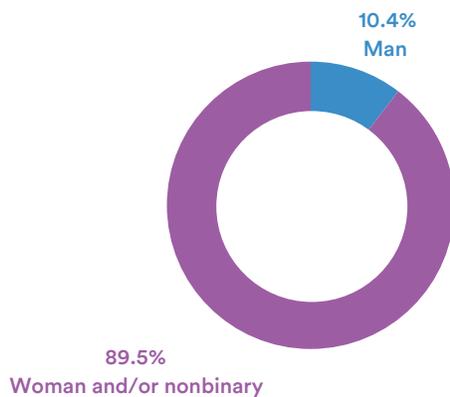


Figure 6.1.5

**PARTICIPANTS of WIML WORKSHOP at NEURIPS (% of TOTAL) by CONTINENT of RESIDENCE, 2020**

Source: Women in Machine Learning, 2020 | Chart: 2021 AI Index Report

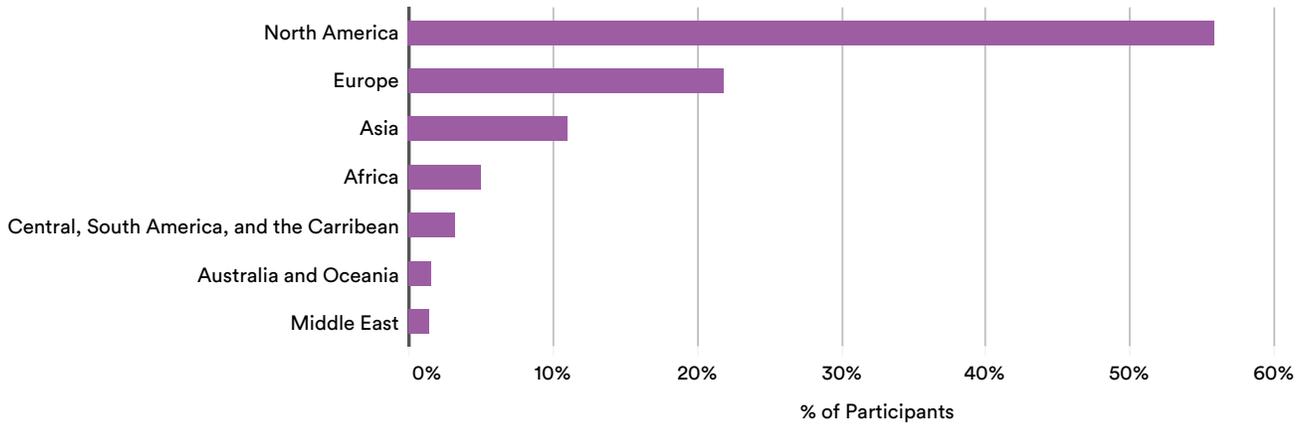


Figure 6.1.6

**PARTICIPANTS of WIML WORKSHOP at NEURIPS (% of TOTAL) by TOP 10 PROFESSIONAL POSITIONS, 2020**

Source: Women in Machine Learning, 2020 | Chart: 2021 AI Index Report

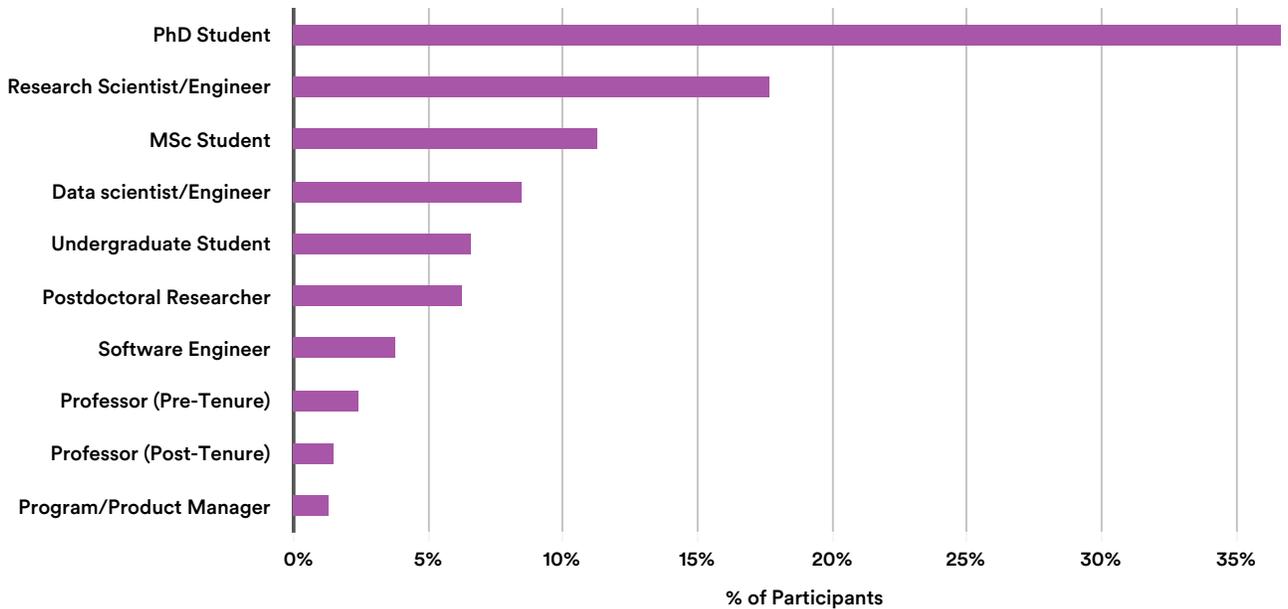


Figure 6.1.7

## 6.2 RACIAL AND ETHNIC DIVERSITY IN AI

### NEW AI PHDS IN THE UNITED STATES BY RACE/ETHNICITY

According to the CRA Taulbee Survey, among the new AI PhDs in 2019 who are U.S. residents, the largest percentage (45.6%) are white (non-Hispanic), followed by Asian (22.4%). By comparison, 2.4% were African American (non-Hispanic) and 3.2% were Hispanic (Figure 6.2.1).

#### NEW U.S. RESIDENT AI PHDS (% of TOTAL) by RACE/ETHNICITY, 2019

Source: CRA Taulbee Survey, 2020 | Chart: 2021 AI Index Report

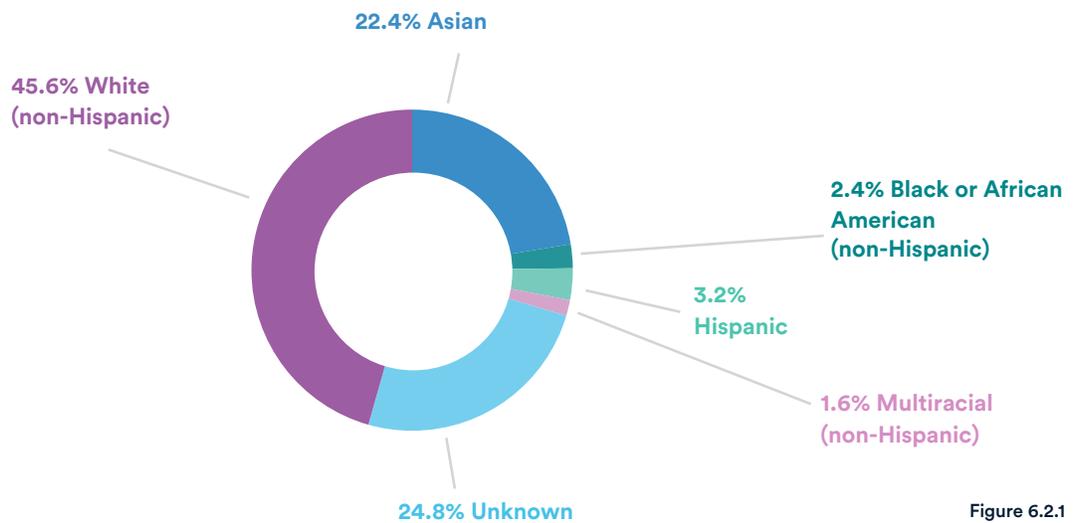


Figure 6.2.1

## NEW COMPUTING PHDS IN THE UNITED STATES BY RACE/ETHNICITY

Figure 6.2.2 shows all PhDs awarded in the United States to U.S. residents across departments of computer science (CS), computer engineering (CE), and information (I) between 2010 and 2019. The CRA survey indicates that the percentage of white (non-Hispanic) new PhDs has changed little over the last 10 years, accounting for 62.7% on average. The share of new Black or African American (non-Hispanic) and Hispanic computing PhDs in the same period is significantly lower, with an average of 3.1% and 3.3%, respectively. We are not able to compare the numbers between new AI and CS PhDs in 2019 because of the number of unknown cases (24.8% for new AI PhDs and 8.5% for CS PhDs).

The CRA survey indicates that the percentage of white (non-Hispanic) new PhDs has changed little over the last 10 years, accounting for 62.7% on average.

### NEW COMPUTING PHDS, U.S. RESIDENT (% of TOTAL) by RACE/ETHNICITY, 2010-19

Source: CRA Taulbee Survey, 2020 | Chart: 2021 AI Index Report

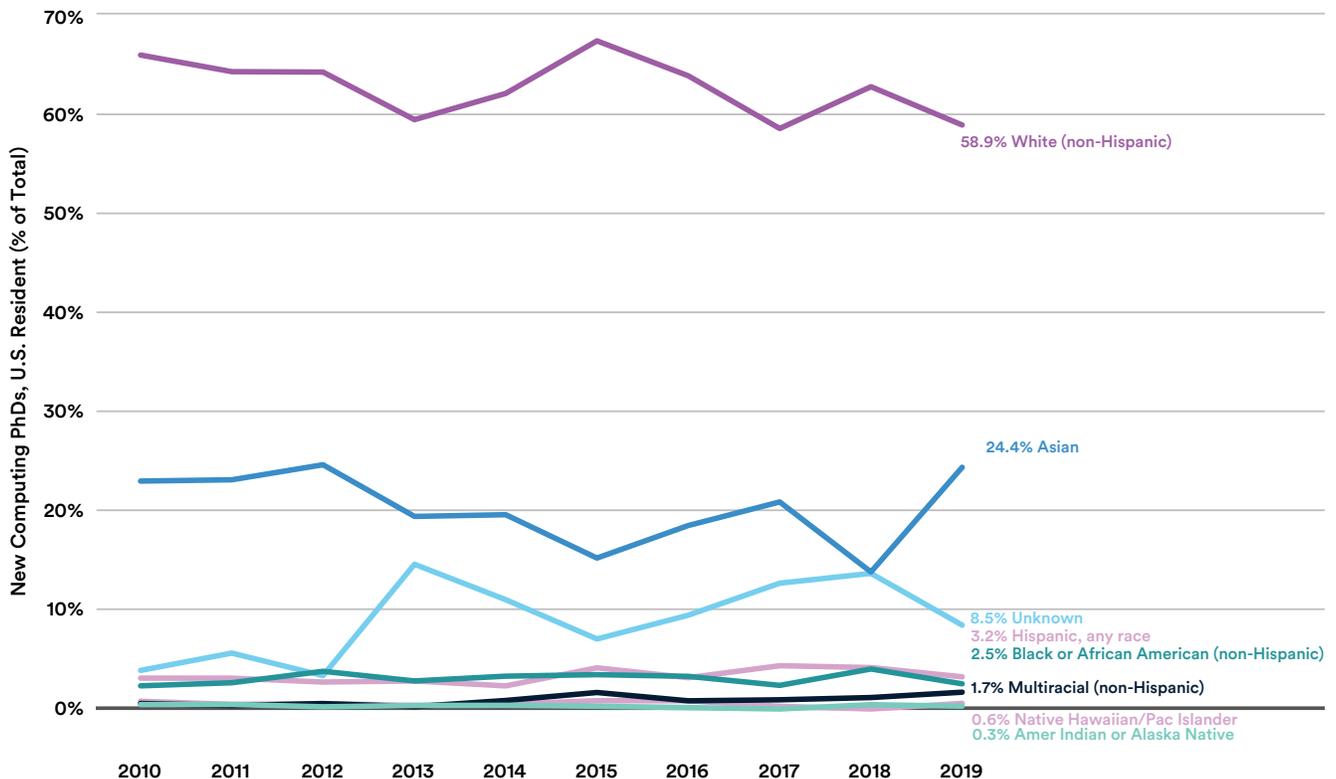


Figure 6.2.2

## CS TENURE-TRACK FACULTY BY RACE/ETHNICITY

Figure 6.2.3 shows data from the AI Index education survey.<sup>3</sup> Among 15 universities that completed the question pertaining to the racial makeup of their faculty, approximately 67.0% of the tenure-track faculty are white, followed by Asian (14.3%), other races (8.3%), and mixed/other race, ethnicity, or origin (6.3%). The smallest representation among tenure-track faculty are teachers of Black or African and of Hispanic, Latino, or Spanish origins, who account for 0.6% and 0.8%, respectively.

**TENURE-TRACK FACULTY (% of TOTAL) at CS DEPARTMENTS of TOP UNIVERSITIES in the WORLD by RACE/ETHNICITY, 2019-20**

Source: AI Index, 2020 | Chart: 2021 AI Index Report

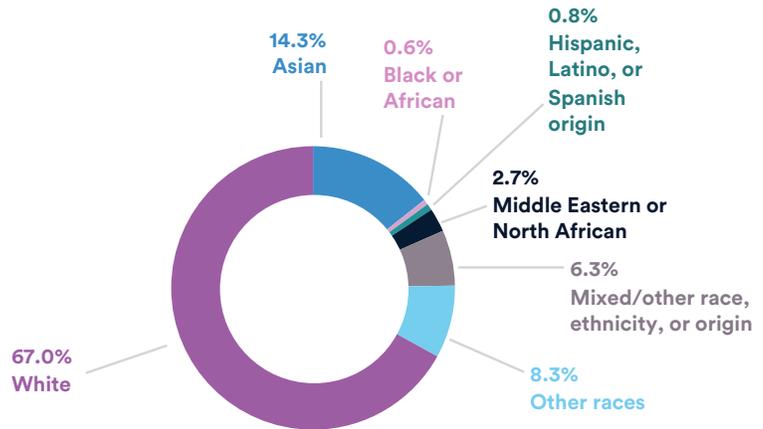


Figure 6.2.3

## BLACK IN AI

**Black in AI (BAI)**, founded in 2017 by Timnit Gebru and Rediet Abebe, is a multi-institutional and transcontinental initiative that aims to increase the presence of Black people in the field of AI. As of 2020, BAI has around 3,000 community members and allies, has held more than 10 workshops at major AI conferences, and has helped increase the number of Black people

participating at major AI conferences globally 40-fold. Figure 6.2.4 shows the number of attendees, submitted papers, and accepted papers from the annual Black in AI Workshop, which is co-located with NeurIPS.<sup>4</sup> The numbers of attendees and accepted papers in 2019 are 2.6 times higher than in 2017, while the number of accepted papers is 2.1 times higher.

**NUMBER OF ATTENDEES, SUBMITTED PAPERS, and ACCEPTED PAPERS at BLACK in AI WORKSHOP CO-LOCATED with NEURIPS, 2017-19**

Source: Black in AI, 2020 | Chart: 2021 AI Index Report

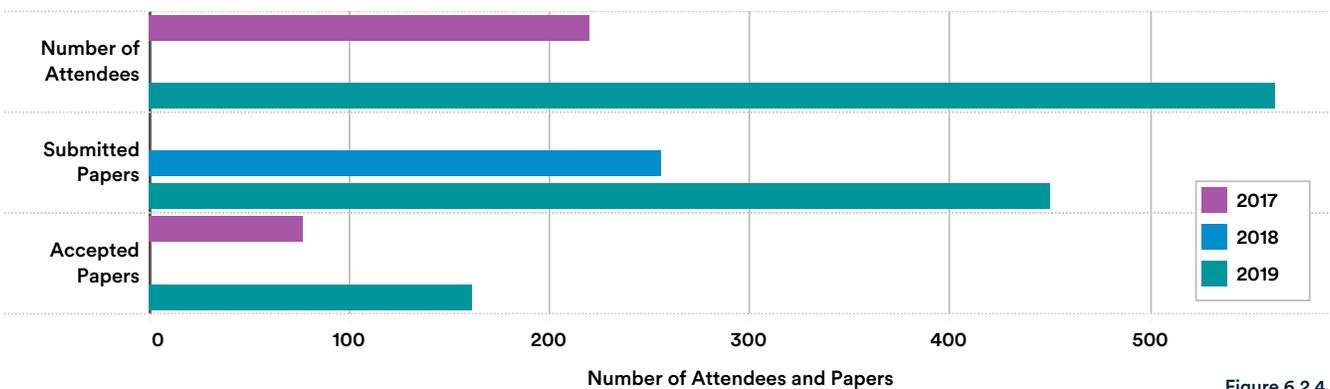


Figure 6.2.4

<sup>3</sup> The survey was distributed to 73 universities online over three waves from November 2020 to January 2021 and completed by 18 universities, a 24.7% response rate. The 18 universities are Belgium: Katholieke Universiteit Leuven; Canada: McGill University; China: Shanghai Jiao Tong University, Tsinghua University; Germany: Ludwig Maximilian University of Munich, Technical University of Munich; Russia: Higher School of Economics, Moscow Institute of Physics and Technology; Switzerland: École Polytechnique Fédérale de Lausanne; United Kingdom: University of Cambridge; United States: California Institute of Technology, Carnegie Mellon University (Department of Machine Learning), Columbia University, Harvard University, Stanford University, University of Wisconsin–Madison, University of Texas at Austin, Yale University.

<sup>4</sup> The 2020 data are clearly affected by the pandemic and not included as a result. For more information, see the [Black in AI impact report](#).

# 6.3 GENDER IDENTITY AND SEXUAL ORIENTATION IN AI

## QUEER IN AI

This section presents data from a membership survey by [Queer in AI \(QAI\)](#),<sup>5</sup> an organization that aims to make the AI/ML community one that welcomes, supports, and values queer scientists. Founded in 2018 by William Agnew, Raphael Gontijo Lopes, and Eva Breznik, QAI builds a visible community of queer and ally AI/ML scientists through meetups, poster sessions, mentoring, and other initiatives.

### Demographics Breakdown

According to the 2020 survey, with around 100 responses,

about 31.5% of respondents identify as gay, followed by bisexual, queer, and lesbian (Figure 6.3.1); around 37.0% and 26.1% of respondents identify as cis male and cis female, respectively, followed by gender queer, gender fluid, nonbinary, and others (Figure 6.3.2). Trans female and male account for 5.0% and 2.5% of total members, respectively. Moreover, the past three years of surveys show that students make up the majority of QAI members—around 41.7% of all respondents on average (Figure 6.3.3), followed by junior-level professionals in academia or industry.

### QAI MEMBERSHIP SURVEY: WHAT IS YOUR SEXUAL ORIENTATION, 2020

Source: Queer in AI, 2020 | Chart: 2021 AI Index Report

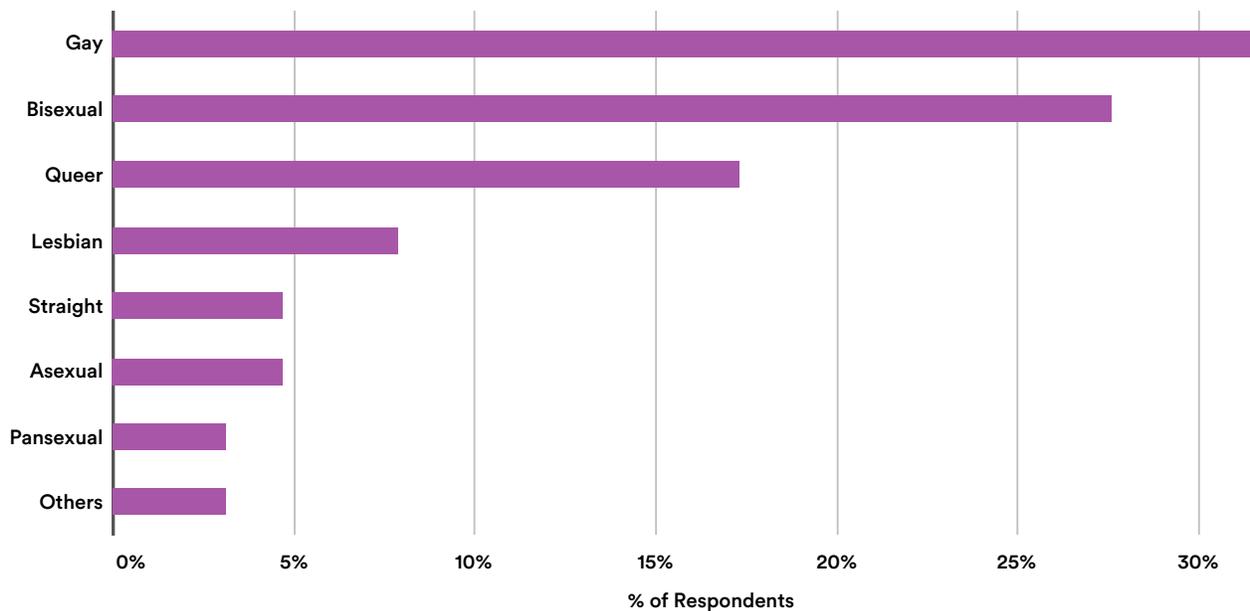


Figure 6.3.1

<sup>5</sup> Queer in AI presents the survey results at its workshop at the annual NeurIPS conference.

### QAI MEMBERSHIP SURVEY: WHAT IS YOUR GENDER IDENTITY, 2020

Source: Queer in AI, 2020 | Chart: 2021 AI Index Report

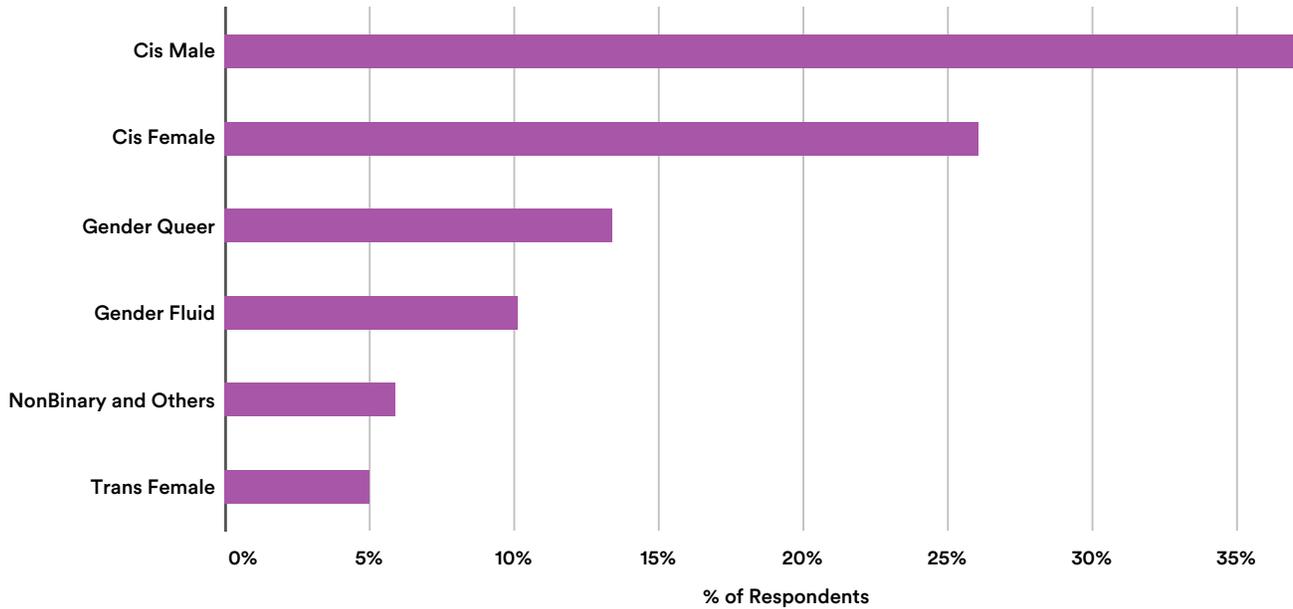


Figure 6.3.2

### QAI MEMBERSHIP SURVEY: HOW WOULD YOU DESCRIBE YOUR POSITION, 2018-20

Source: Queer in AI, 2020 | Chart: 2021 AI Index Report

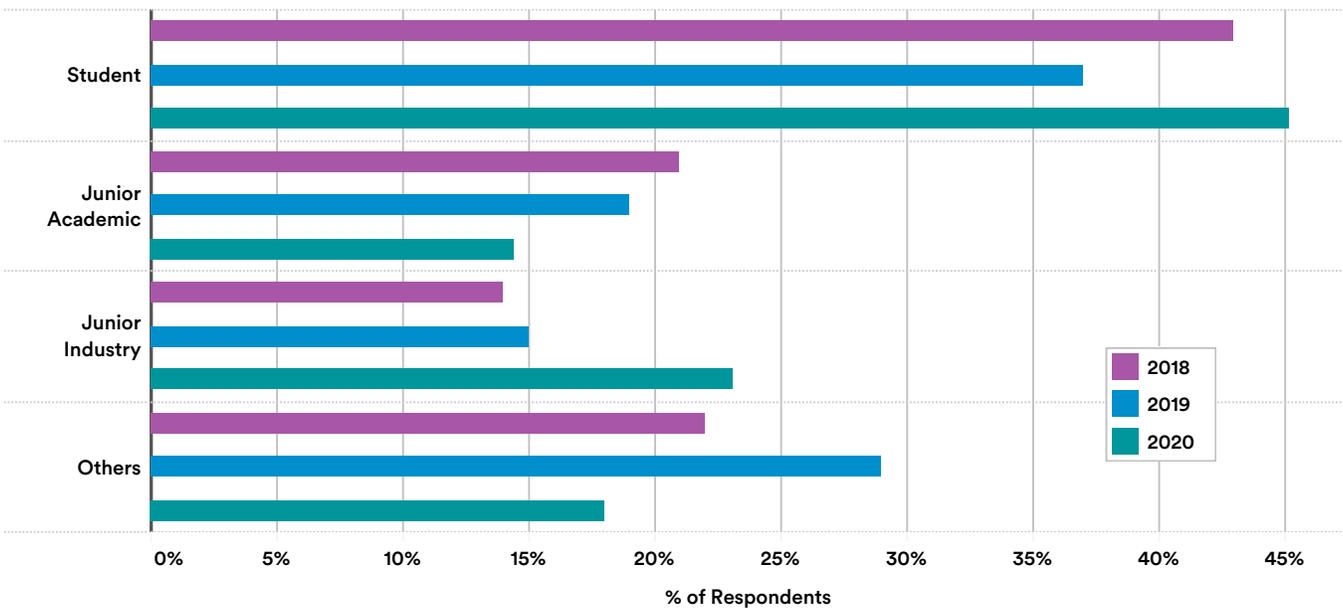


Figure 6.3.3

### Experience as Queer Practitioners

QAI also surveyed its members on their experiences as queer AI/ML practitioners. As shown in Figure 6.3.4, 81.4% regard the lack of role models as being a major obstacle for their careers, and 70.9% think the lack of community contributes to the same phenomenon. Almost half the respondents also view the lack of inclusiveness in the field as an obstacle. Moreover, more than 40% of QAI members have experienced discrimination or harassment as a queer person at work or school (Figure 6.3.5). Around 9.7% have encountered discrimination or harassment on more than five occasions.

**Among surveyed QAI members, 81.4% regard the lack of role models as being a major obstacle for their careers, and 70.9% think the lack of community contributes to the same phenomenon.**

#### QAI MEMBERSHIP SURVEY: WHAT ARE OBSTACLES YOU HAVE FACED in BECOMING a QUEER AI/ML PRACTITIONER, 2020

Source: Queer in AI, 2020 | Chart: 2021 AI Index Report

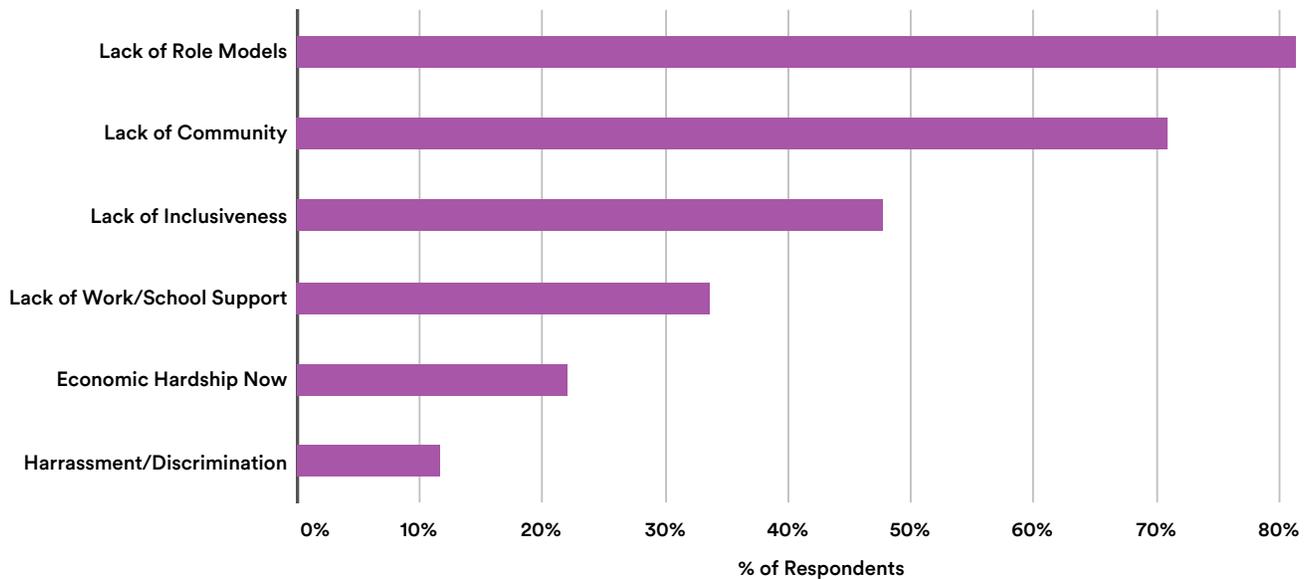


Figure 6.3.4

More than 40% of QAI members have experienced discrimination or harassment as a queer person at work or school. Around 9.7% have encountered discrimination or harassment on more than five occasions.

**QAI MEMBERSHIP SURVEY: HAVE YOU EXPERIENCED DISCRIMINATION/HARASSMENT as a QUEER PERSON at YOUR JOB or SCHOOL, 2020**

Source: Queer in AI, 2020 | Chart: 2021 AI Index Report

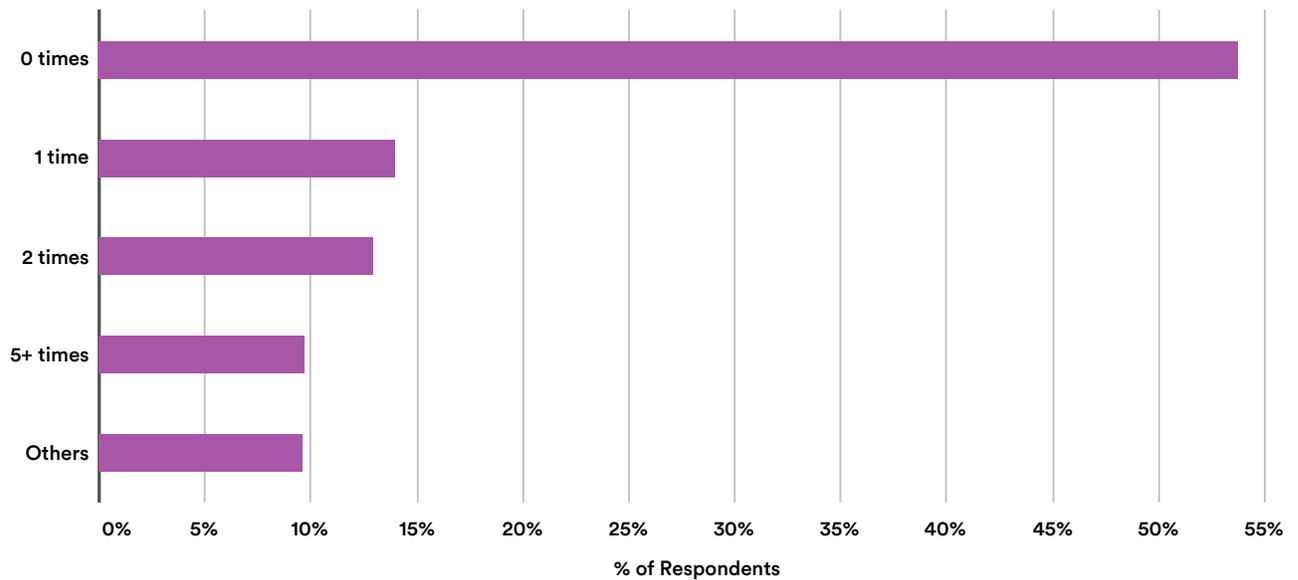


Figure 6.3.5



# APPENDIX

## LINKEDIN

### AI Skills Penetration

The aim of this indicator is to measure the intensity of AI skills in an entity (in a particular country, industry, gender, etc.) through the following methodology:

- Compute frequencies for all self-added skills by LinkedIn members in a given entity (occupation, industry, etc.) in 2015–2020.
- Re-weight skill frequencies using a TF-IDF model to get the top 50 most representative skills in that entity. These 50 skills compose the “skill genome” of that entity.
- Compute the share of skills that belong to the AI skill group out of the top skills in the selected entity.

**Interpretation:** The AI skill penetration rate signals the prevalence of AI skills across occupations, or the intensity with which LinkedIn members utilize AI skills in their jobs. For example, the top 50 skills for the occupation of engineer are calculated based on the weighted frequency with which they appear in LinkedIn members’ profiles. If four of the skills that engineers possess belong to the AI skill group, this measure indicates that the penetration of AI skills is estimated to be 8% among engineers (e.g., 4/50).

### Relative AI Skills Penetration

To allow for skills penetration comparisons across countries, the skills genomes are calculated and a relevant benchmark is selected (e.g., global average). A ratio is then constructed between a country’s and the benchmark’s AI skills penetrations, controlling for occupations.

**Interpretation:** A country’s relative AI skills penetration of 1.5 indicates that AI skills are 1.5 times as frequent as in the benchmark, for an overlapping set of occupations.

### Global Comparison: By Gender

The relative AI skill penetration by country for gender provides an in-depth decomposition of AI skills penetration across female and male labor pools and sample countries.

**Interpretation:** A country’s relative AI skill penetration rate of 2 for women means that the average penetration of AI skills among women in that country is two times the global average across the same set of occupations among women. If, in the same country, the relative AI skill penetration rate for men is 1.9, this indicates that the average penetration of AI skills among women in that country is 5% higher than that of men (calculated by dividing 1.9 by 2 and then subtracting 1, or  $2/1.9-1$ ) for the same set of occupations.