Funded and unfunded science in Russia: A new dataset and longitudinal analysis

Katie Spoon, Elena Chechik, Vinicius Muraro and Xinzhe Li

ATLC, May 15, 2025

katiespoon.github.io/
kspoon.bsky.social
katherine.spoon@colorado.edu

Basic scientific discoveries are enabled by governmental funding in research. Example: the lizard venom study that led to Ozempic!

Early 1990s

VA researchers in the US discovered a compound in Gila venom that stimulates the body's insulin production and licensed it to a pharmaceutical company for drug development.

Basic scientific discoveries are enabled by governmental funding in research. Example: the lizard venom study that led to Ozempic!

Early 1990s

VA researchers in the US discovered a compound in Gila venom that stimulates the body's insulin production and licensed it to a pharmaceutical company for drug development.

After more R&D, the discovery eventually led to GLP-1 drugs including Ozempic, the diabetes and weight-loss medicine.

Basic scientific discoveries are enabled by governmental funding in research. Example: the lizard venom study that led to Ozempic!

Early 1990s

VA researchers in the US discovered a compound in Gila venom that stimulates the body's insulin production and licensed it to a pharmaceutical company for drug development.

After more R&D, the discovery eventually led to GLP-1 drugs including Ozempic, the diabetes and weight-loss medicine.

These drugs have improved & will improve the lives of many people

Basic scientific discoveries are enabled by governmental funding in research. Example: the lizard venom study that led to Ozempic!

Basic scientific discoveries are enabled by governmental funding in research. Example: the lizard venom study that led to Ozempic!

[Photos: Munro/Getty Images, Milan Zygmunt/Getty Images, Unsplash]
Fast Company, https://www.fastcompany.com/91276813/ozempic-research-nih-cuts
https://www.research.va.gov/research_in_action/Diabetes-drug-from-Gila-monster-venom.cfm

Which topics are *not* funded?

While funding is a critical driver of scientific progress, it remains unknown whether and to what extent certain topics are unfunded.

Which topics are *not* funded?

While funding is a critical driver of scientific progress, it remains unknown whether and to what extent certain topics are unfunded.

Research is limited because data about ***unfunded*** grant proposals is difficult to get access to.

Which topics are *not* funded?

While funding is a critical driver of scientific progress, it remains unknown whether and to what extent certain topics are unfunded.

Research is limited because data about ***unfunded*** grant proposals is difficult to get access to.

Studies that analyze unfunded proposals focus on the U.S. and tend to analyze a single cross-section in time, rather than comparing how unfunded topics at an agency change over time.

Taffe & Gilpin. (2021). Racial inequity in grant funding from the US National Institutes of Health. Elife, 10, e65697.
Horbach, Tijdink & Bouter. (2022). Research funders should be more transparent: A plea for open applications. Royal Society Open Science, 9(10), 220750.

326,661 supported and rejected grant applications

RUSSIAN FOUNDATION FOR BASIC RESEARCH
RFBR

RUSSIAN FOUNDATION FOR BASIC RESEARCH

326,661 supported and rejected grant applications

22 years, 1994-2016

Eight disciplines (Biology & Medical Sciences, Chemistry & Material Sciences, Earth Sciences, Engineering, Humanities & Social Sciences, Information Technology, Math & Computer Science, and Physics & Astronomy)

We constructed the dataset using web-scraping and Russian language assistance from an expert. We scraped each submitted project to the RFBR including the year, title, field, competition, and status (accepted or rejected). We removed duplicates, projects with missing data, and competitions that did not directly fund research (e.g., a competition for conference participation).

RUSSIAN FOUNDATION FOR BASIC RESEARCH

RFBR

326,661 supported and rejected grant applications

22 years, 1994-2016

Eight disciplines (Biology & Medical Sciences, Chemistry & Material Sciences, Earth Sciences, Engineering, Humanities & Social Sciences, Information Technology, Math & Computer Science, and Physics & Astronomy)

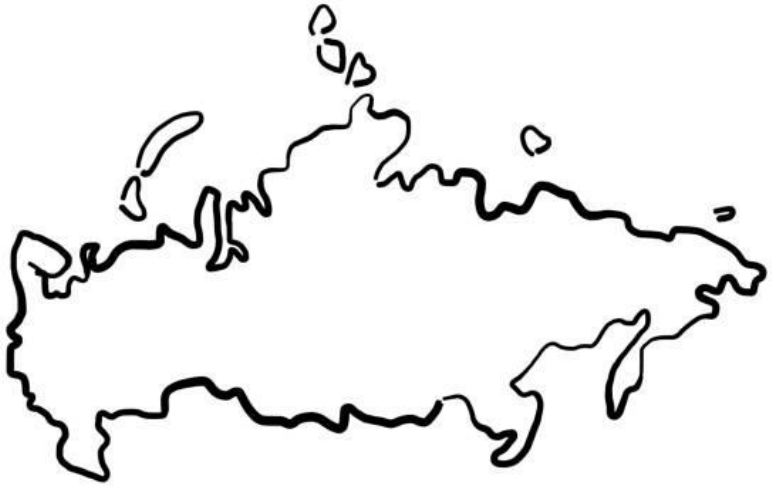
Annotated with estimates of gender and academic experience

Are the topics that are historically under-funded those that women disproportionately study?

We estimated the gender of each applicant using algorithmic name-based gender associations. We estimated the academic experience of each applicant from information about the funding competitions by age (max) and number of publications (min). For example, one early-career competition required applicants to be under age 35 and have more than two recent publications. Our dataset has suitable statistical power to assess differences between subgroups (e.g., women in a given field and year who were funded versus men).

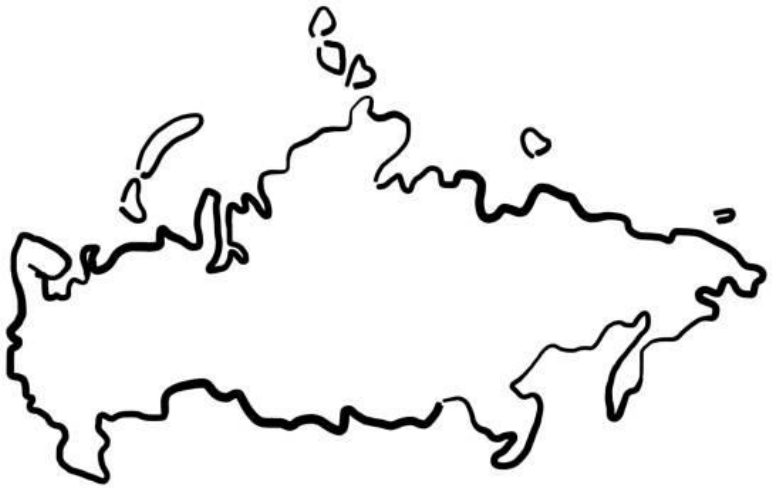
We constructed the dataset using web-scraping and Russian language assistance from an expert. We scraped each submitted project to the RFBR including the year, title, field, competition, and status (accepted or rejected). We removed duplicates, projects with missing data, and competitions that did not directly fund research (e.g., a competition for conference participation).

Why Russia?



Identifiable sociopolitical changes over time

Why Russia?

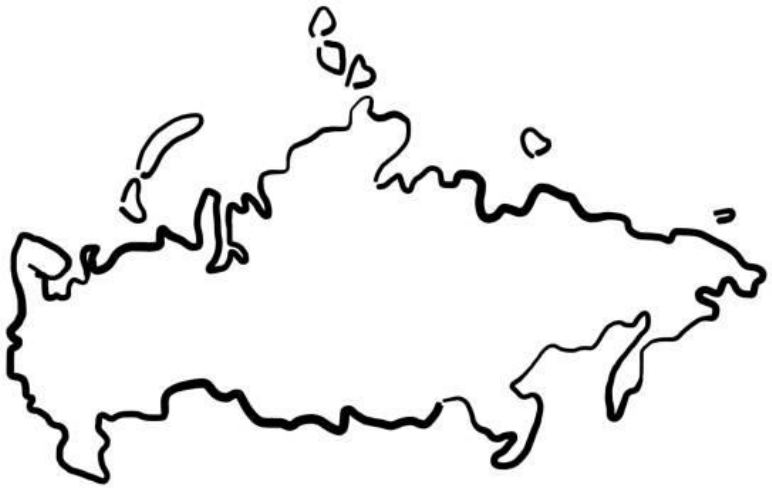


Identifiable sociopolitical changes over time

Weak state control, low funding

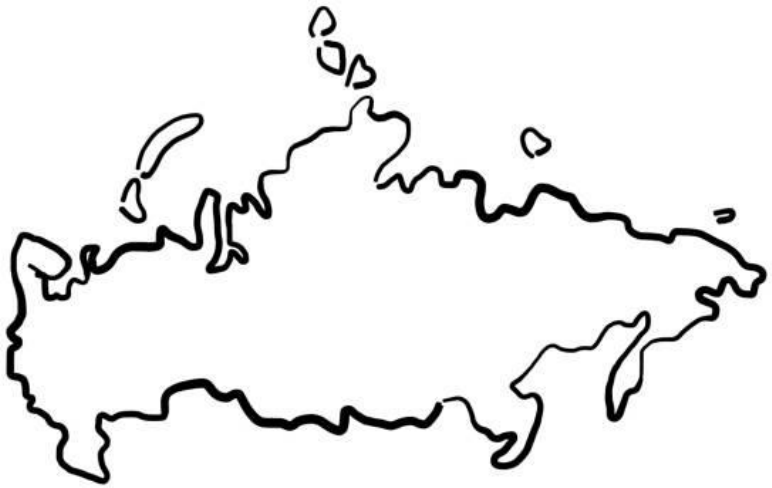
1994 2004

Why Russia?



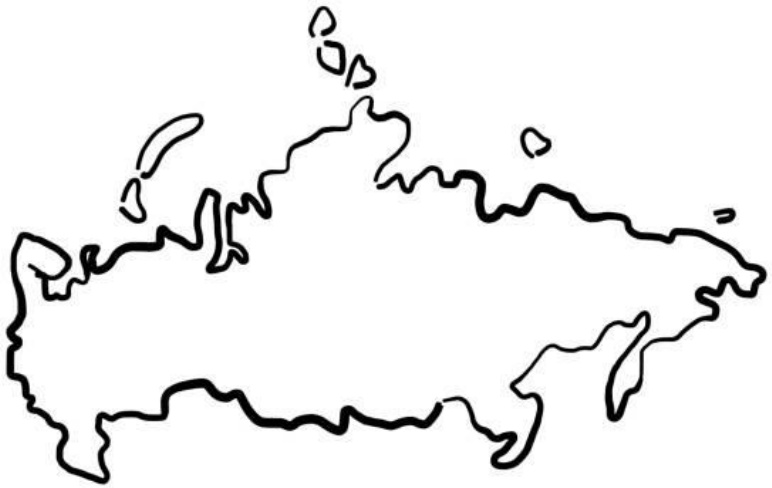
Identifiable sociopolitical changes over time

Why Russia?



Identifiable sociopolitical changes over time

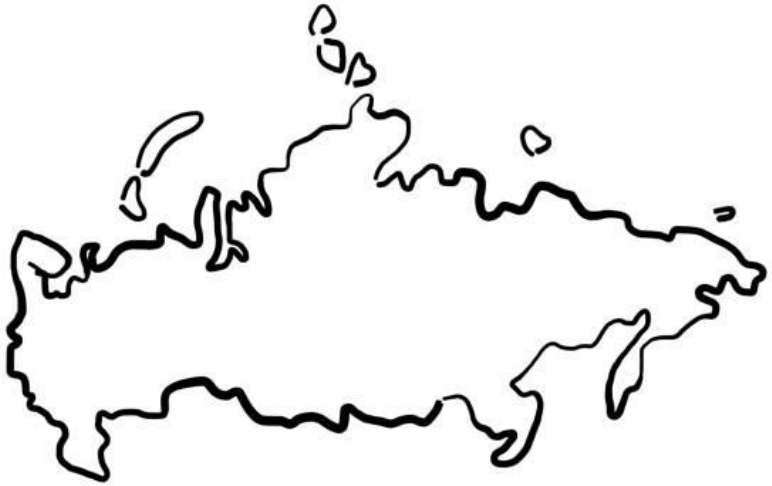
Why Russia?



Identifiable sociopolitical changes over time

…allow us to study how politics may influence scientific funding and understand the political nature of scientific funding worldwide.

Why Russia?



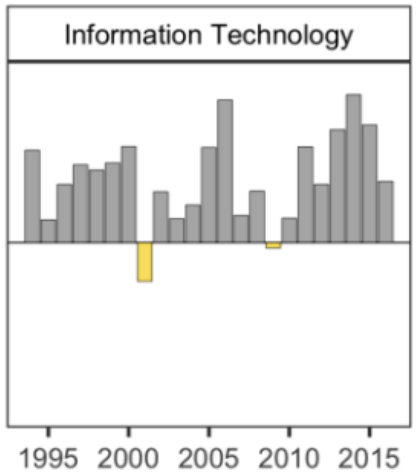
Also, gender dynamics! Strong participation of women in science & gender detection in Russian names is very straightforward.

Identifiable sociopolitical changes over time

…allow us to study how politics may influence scientific funding and understand the political nature of scientific funding worldwide.

Are the topics that are historically under-funded those that women disproportionately study?

Are the topics that are historically under-funded those that women disproportionately study?



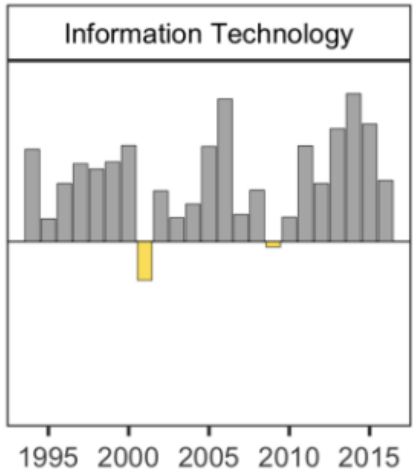
Makes up the smallest share of proposals

But most rapidly growing share over time

Are the topics that are historically under-funded those that women disproportionately study?

gender gap in favor of men gender gap in favor of women

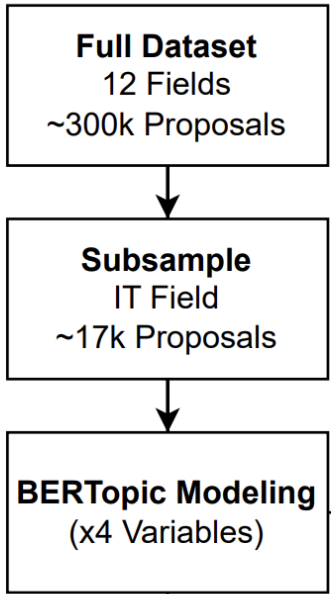
Discrimination or topics?

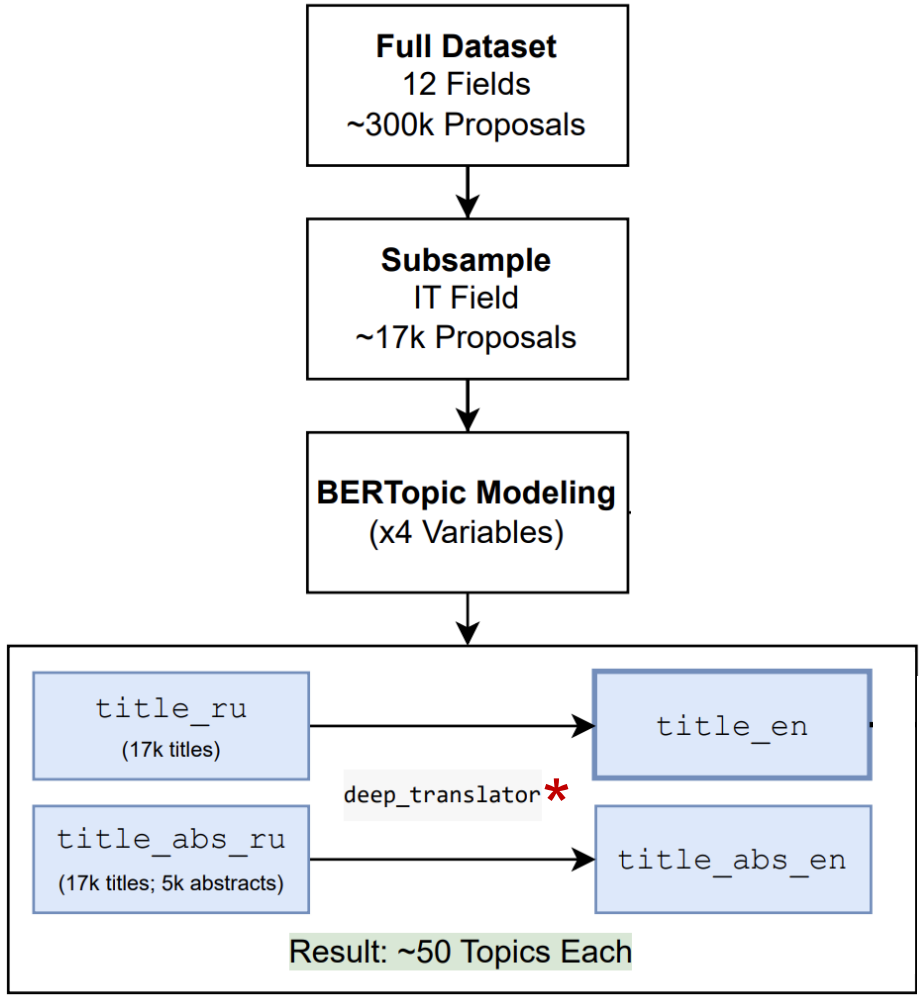


Makes up the smallest share of proposals

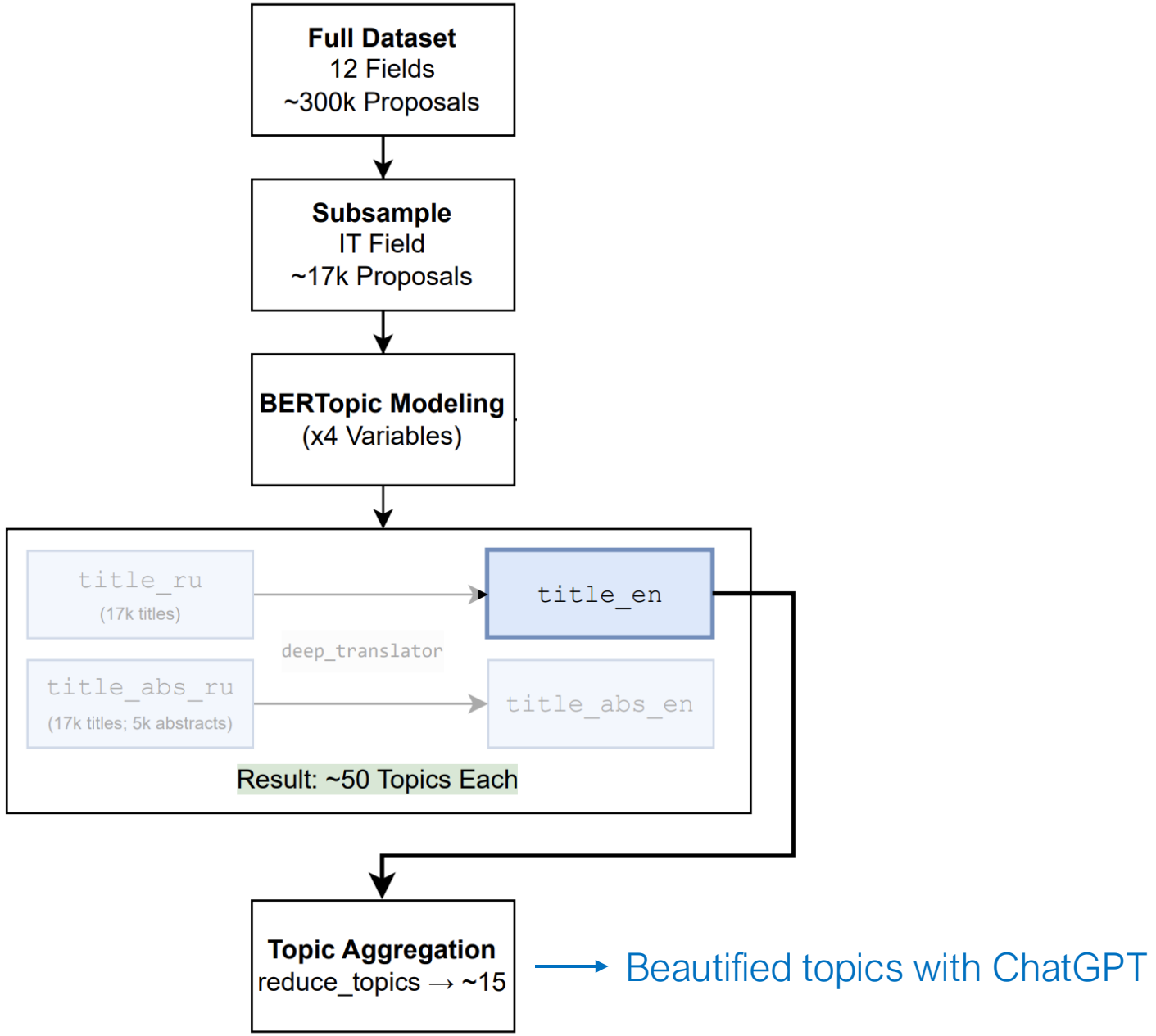
But most rapidly growing share over time

Steep gender gaps in favor of men among recipients


Full Dataset
12 Fields
~300k Proposals
Subsample
IT Field
~17k Proposals
BERTopic Modeling
(x4 Variables)

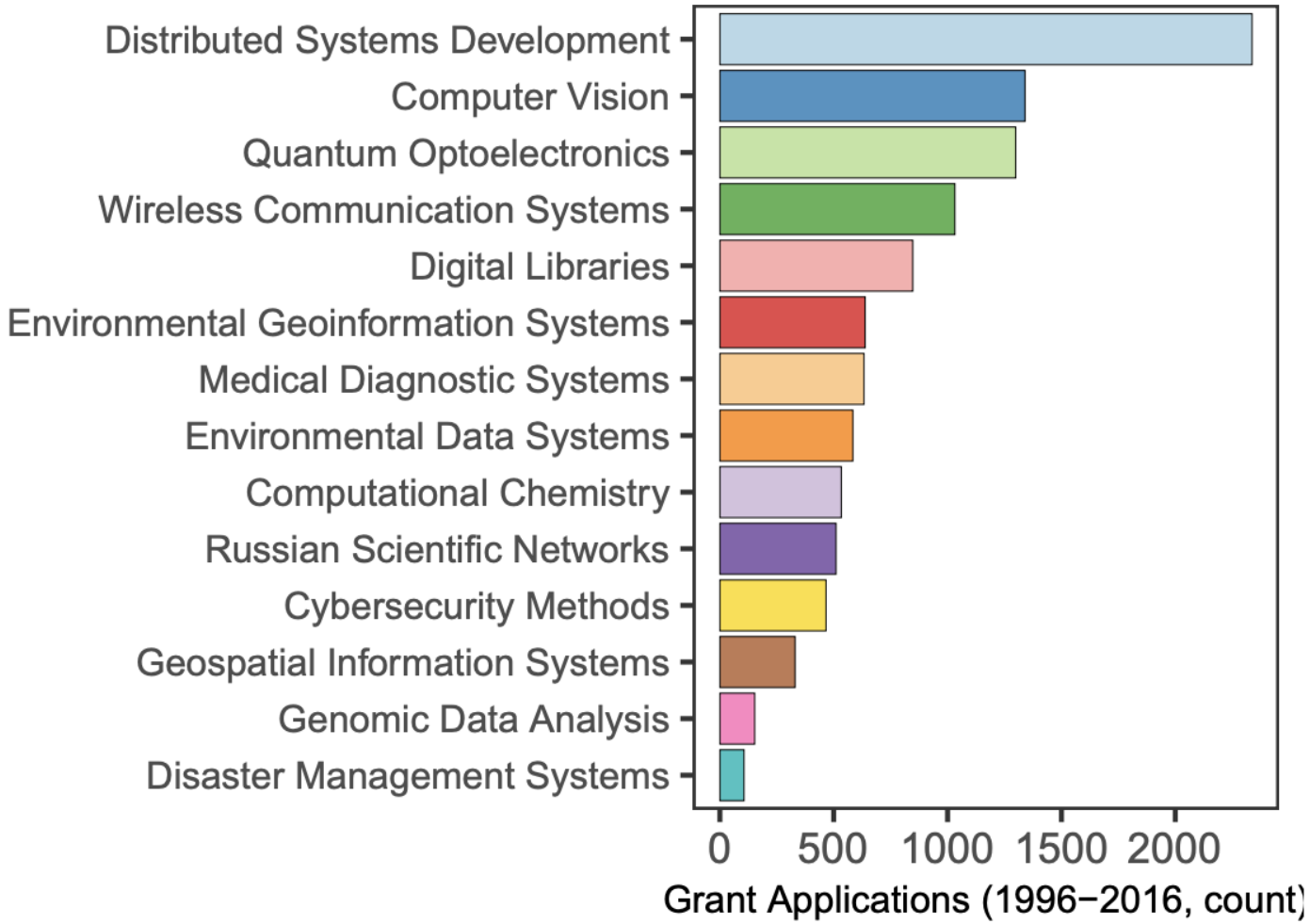


Currently our bottleneck, very slow

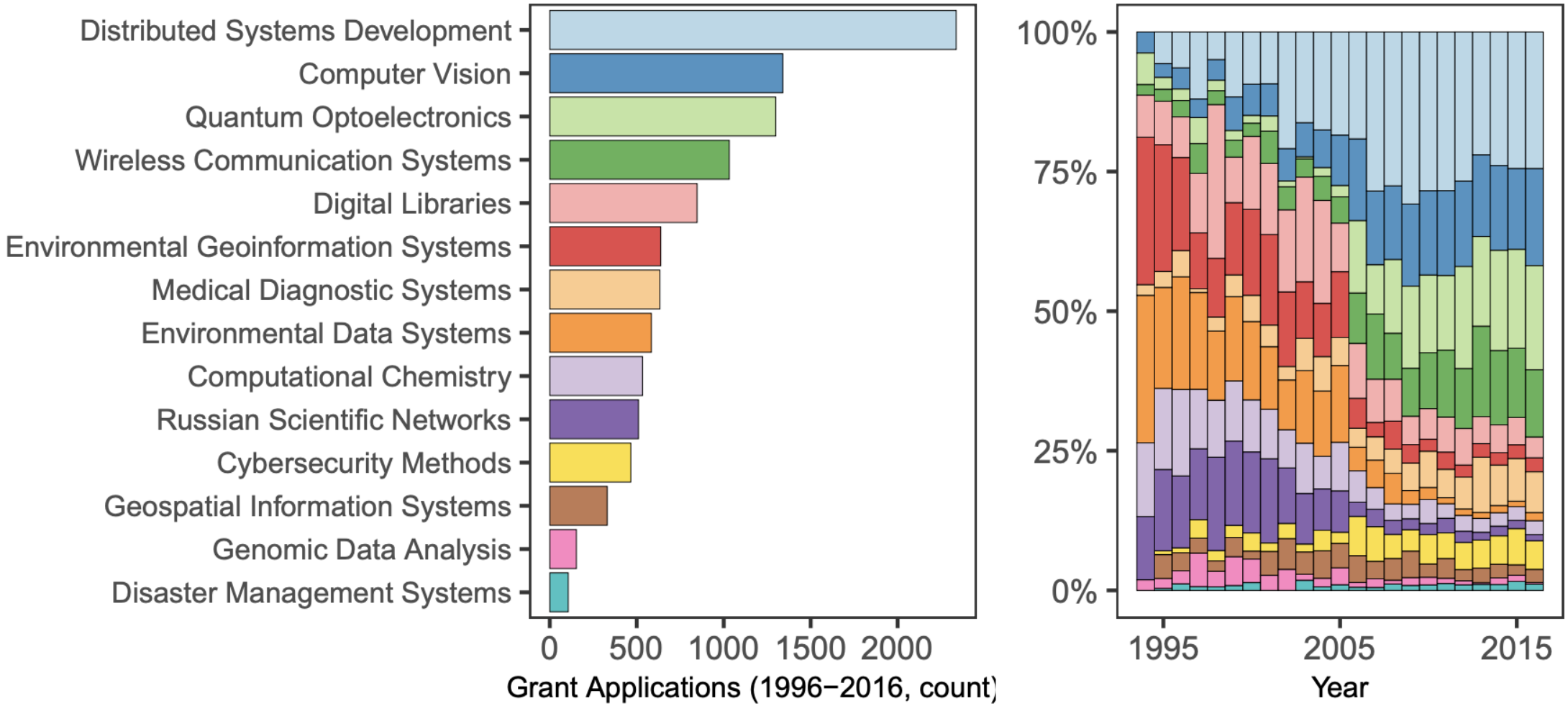


Kozlowski, Pradier, & Benz. (2024). Generative AI for automatic topic labelling. arXiv preprint arXiv:2408.07003.

Topics in Information Technology

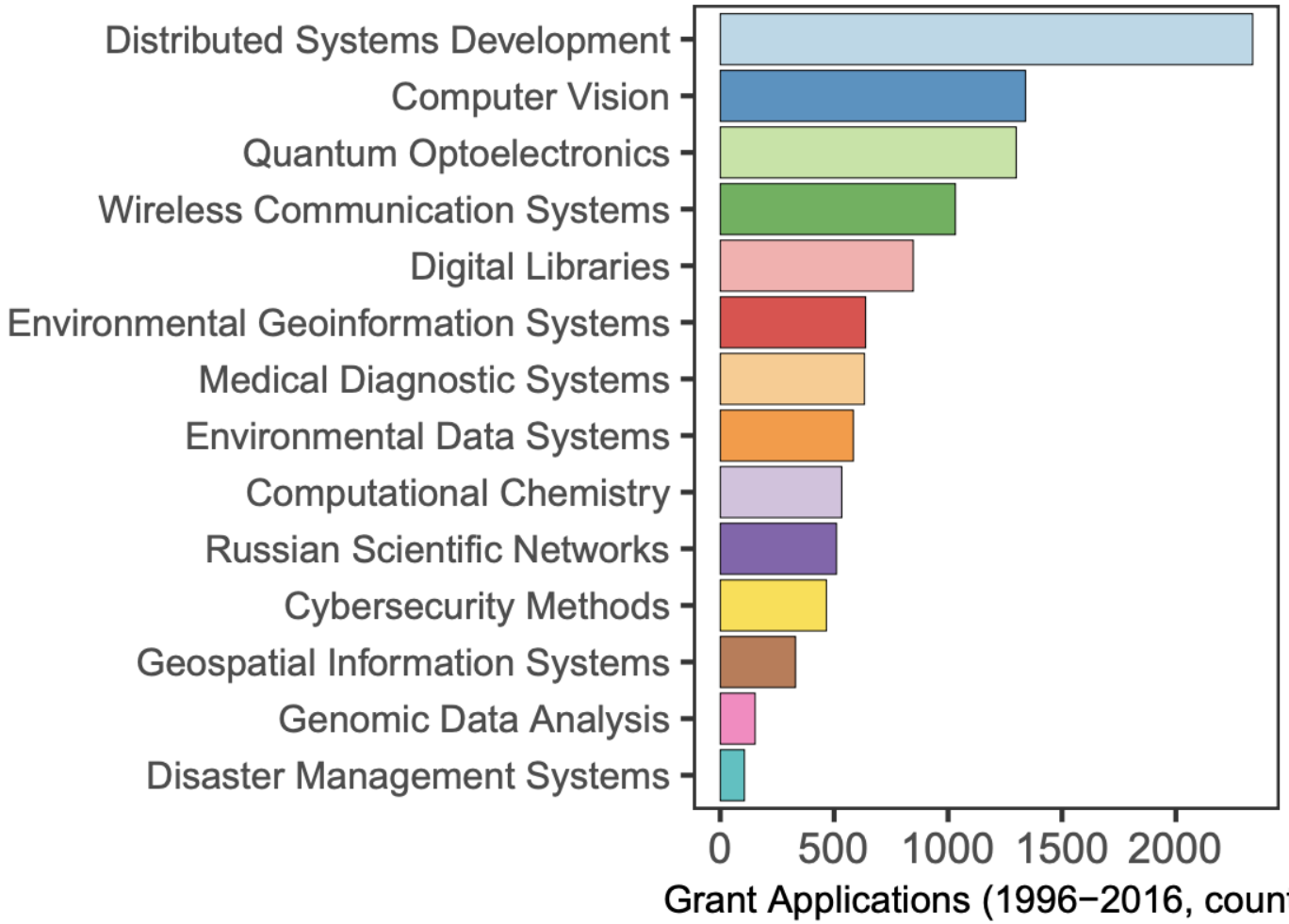


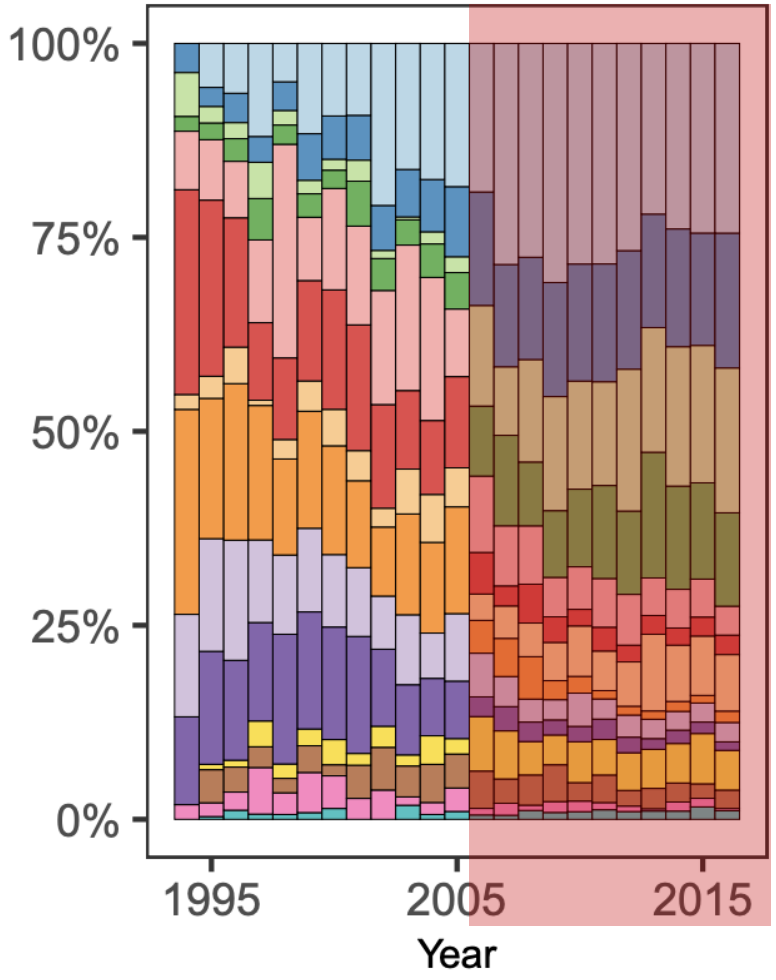
Topics in Information Technology



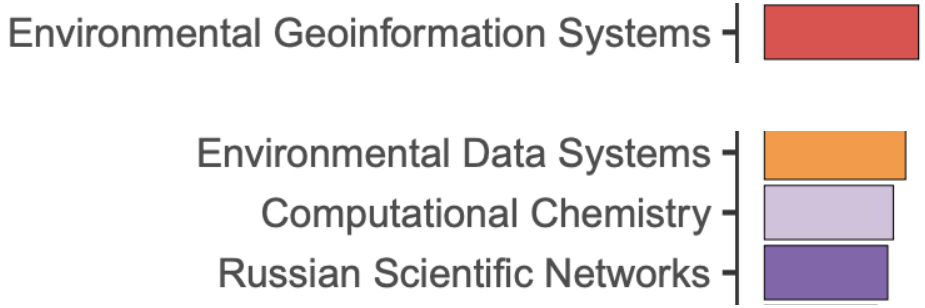
Topics in Information Technology

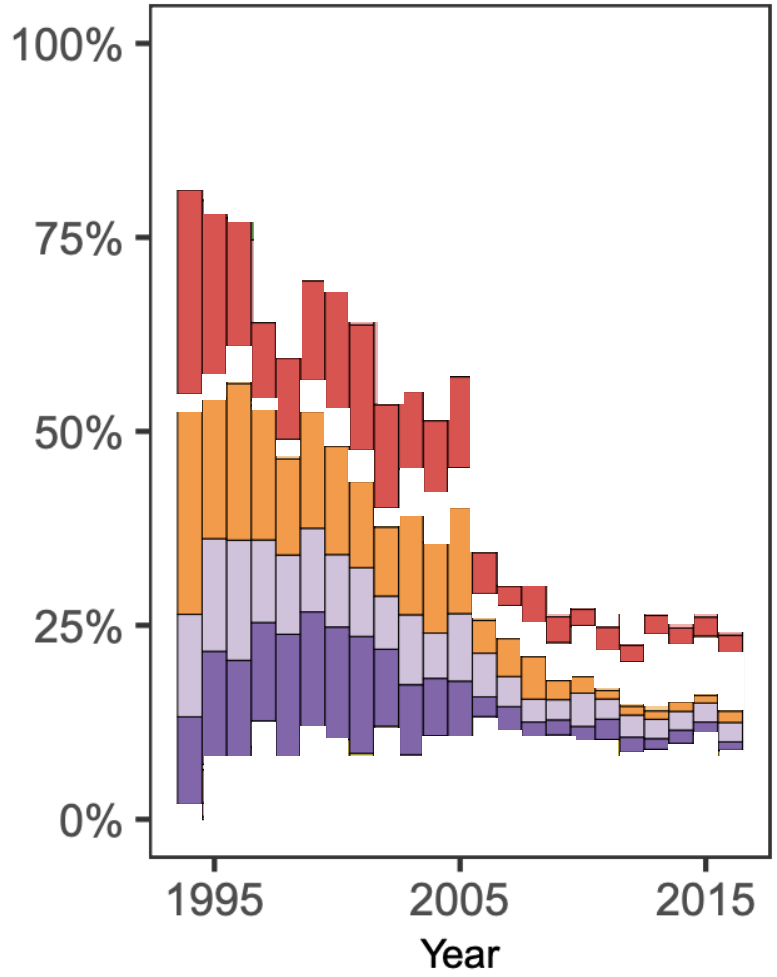
Big economic developments
Increasing state control



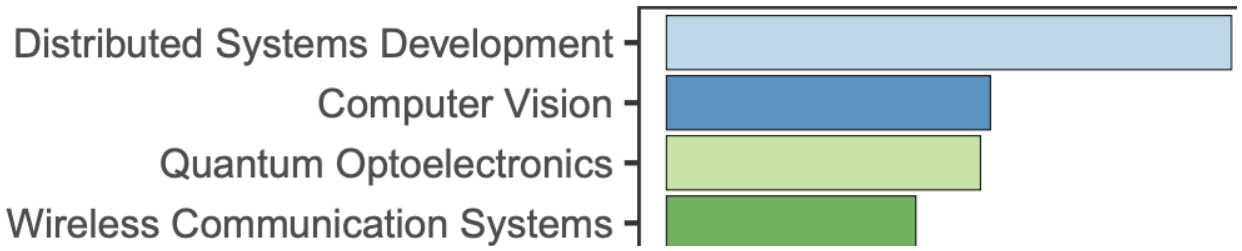


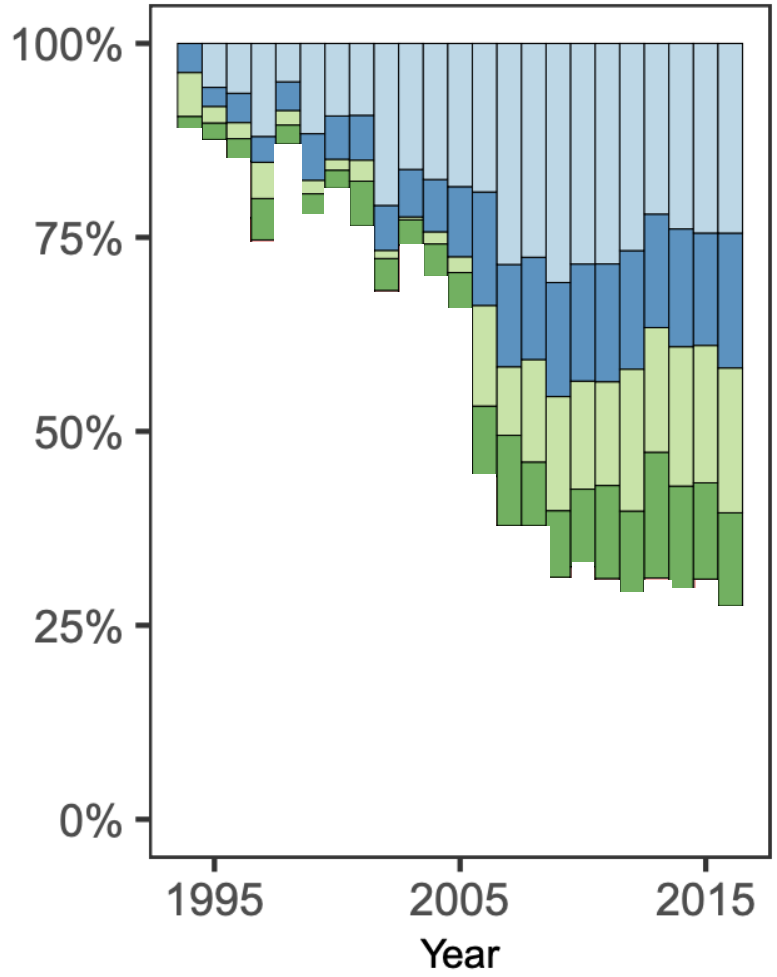
Topics in Information Technology



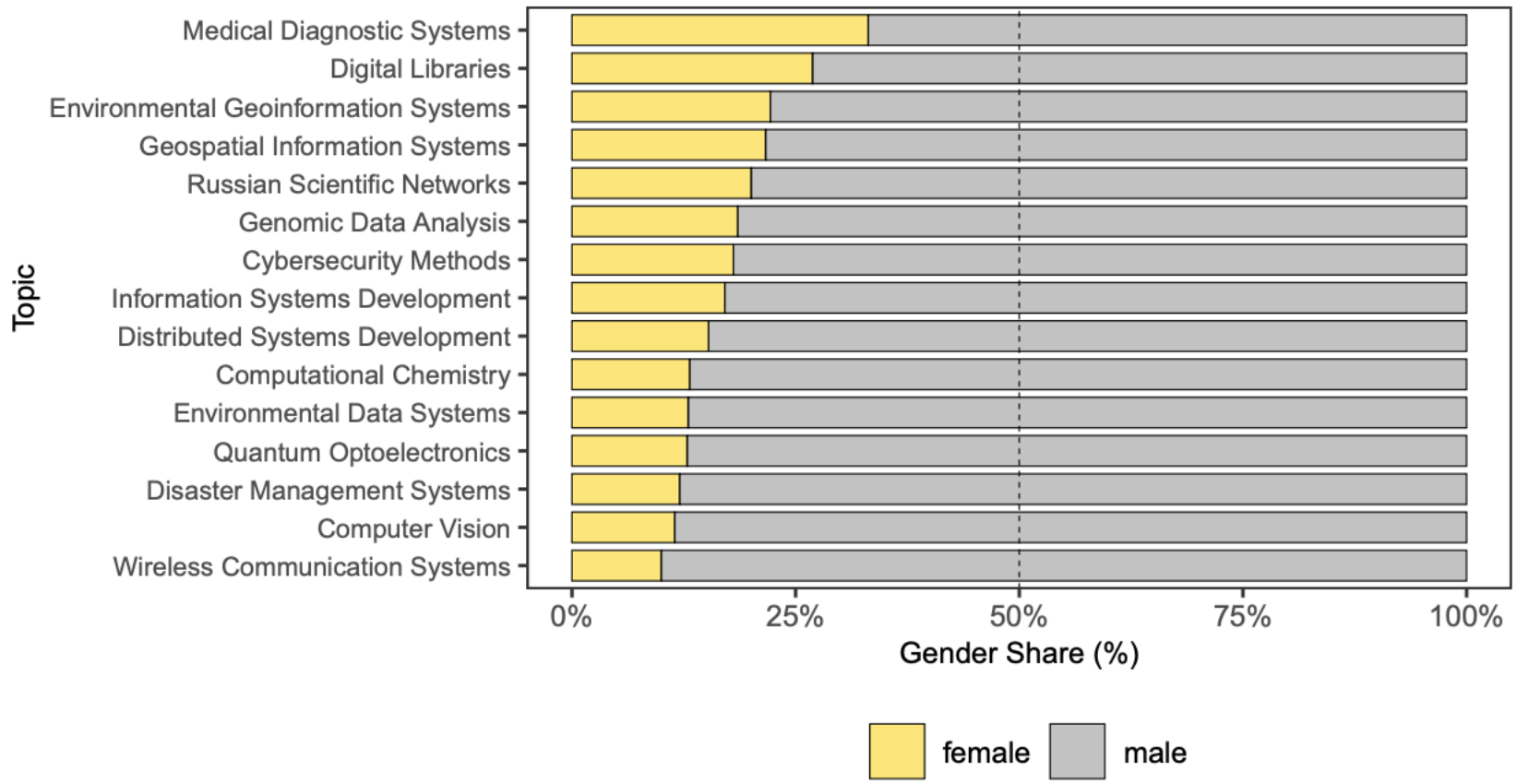


Topics in Information Technology

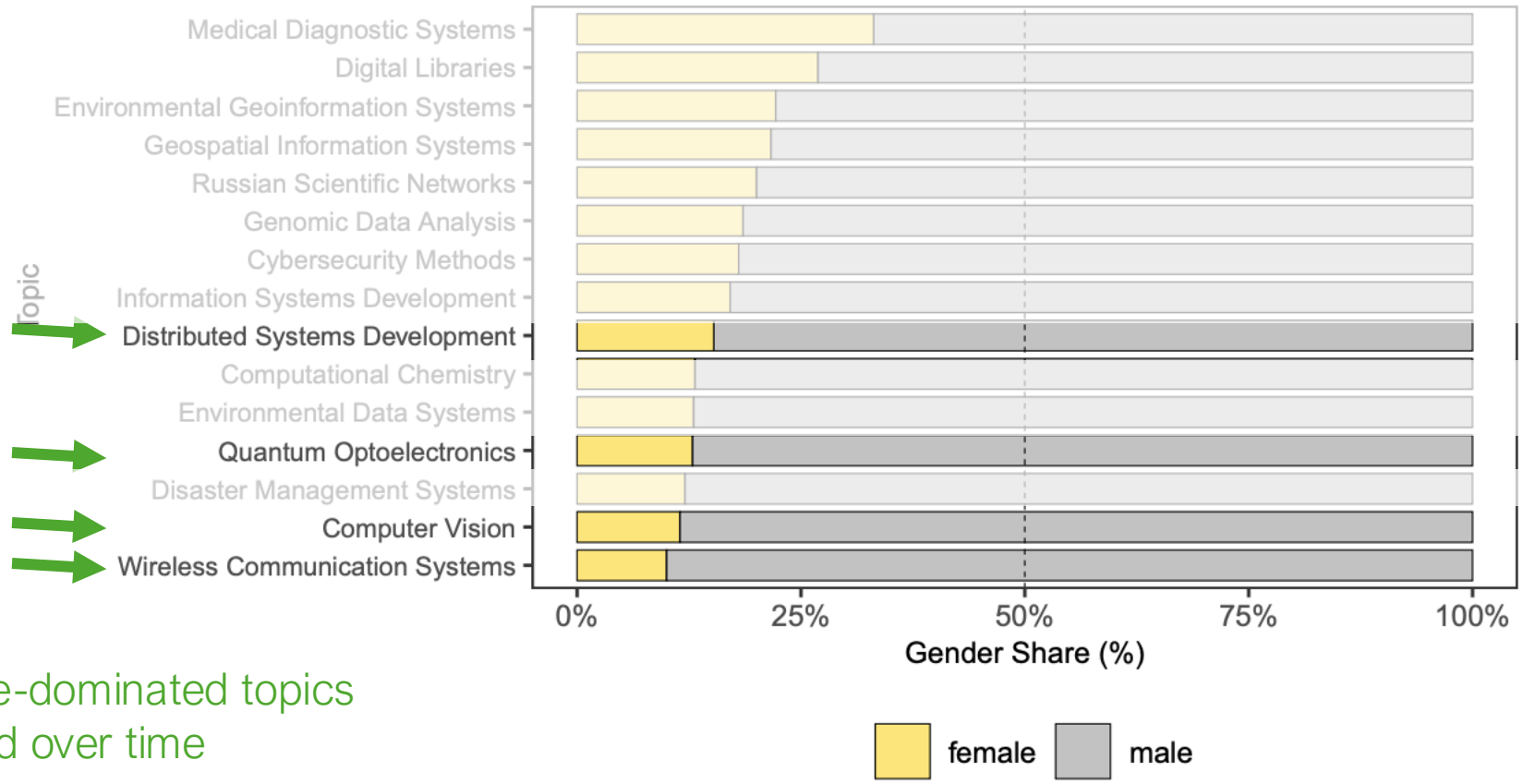




Share of applications by topic & gender

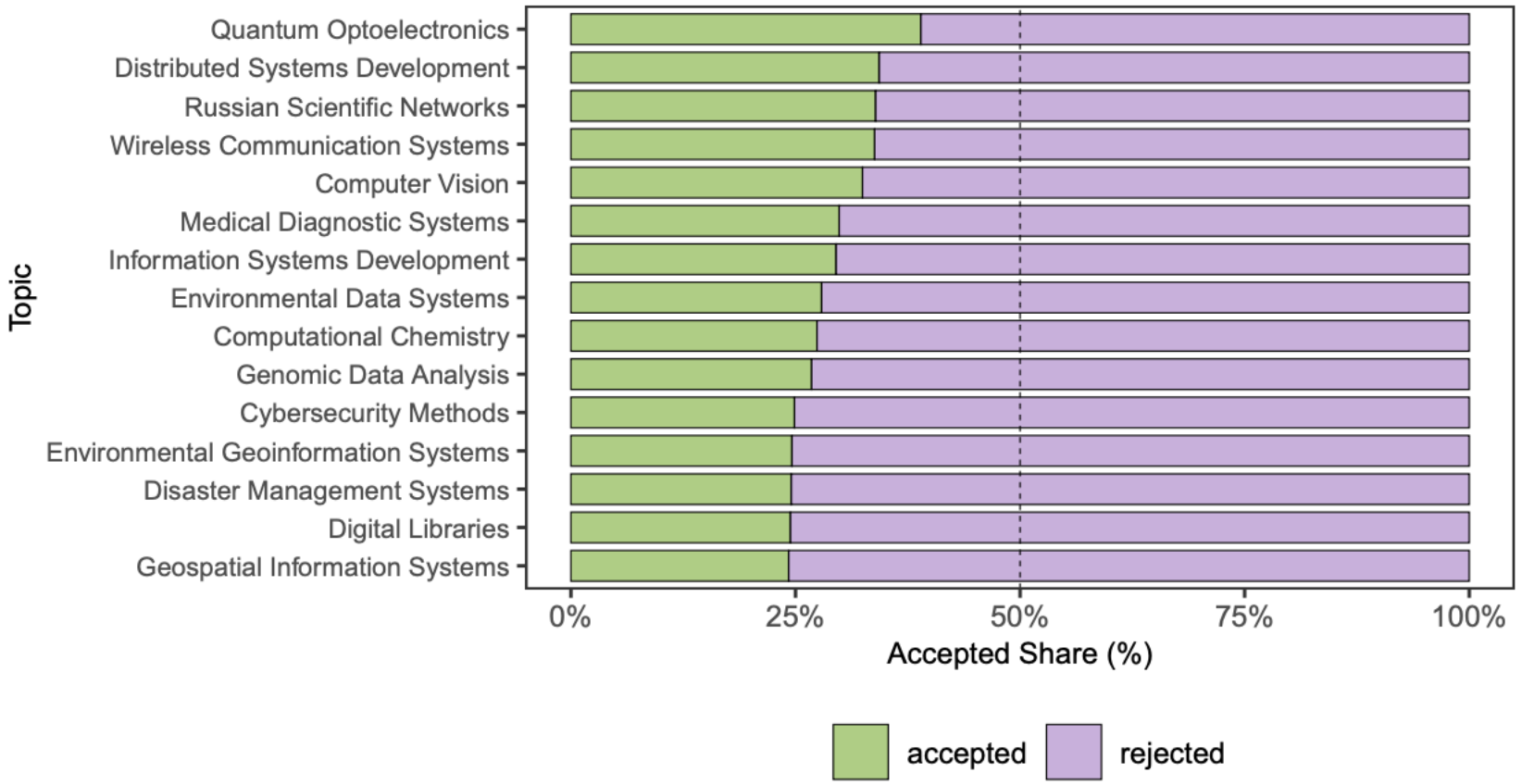


Share of applications by topic & gender

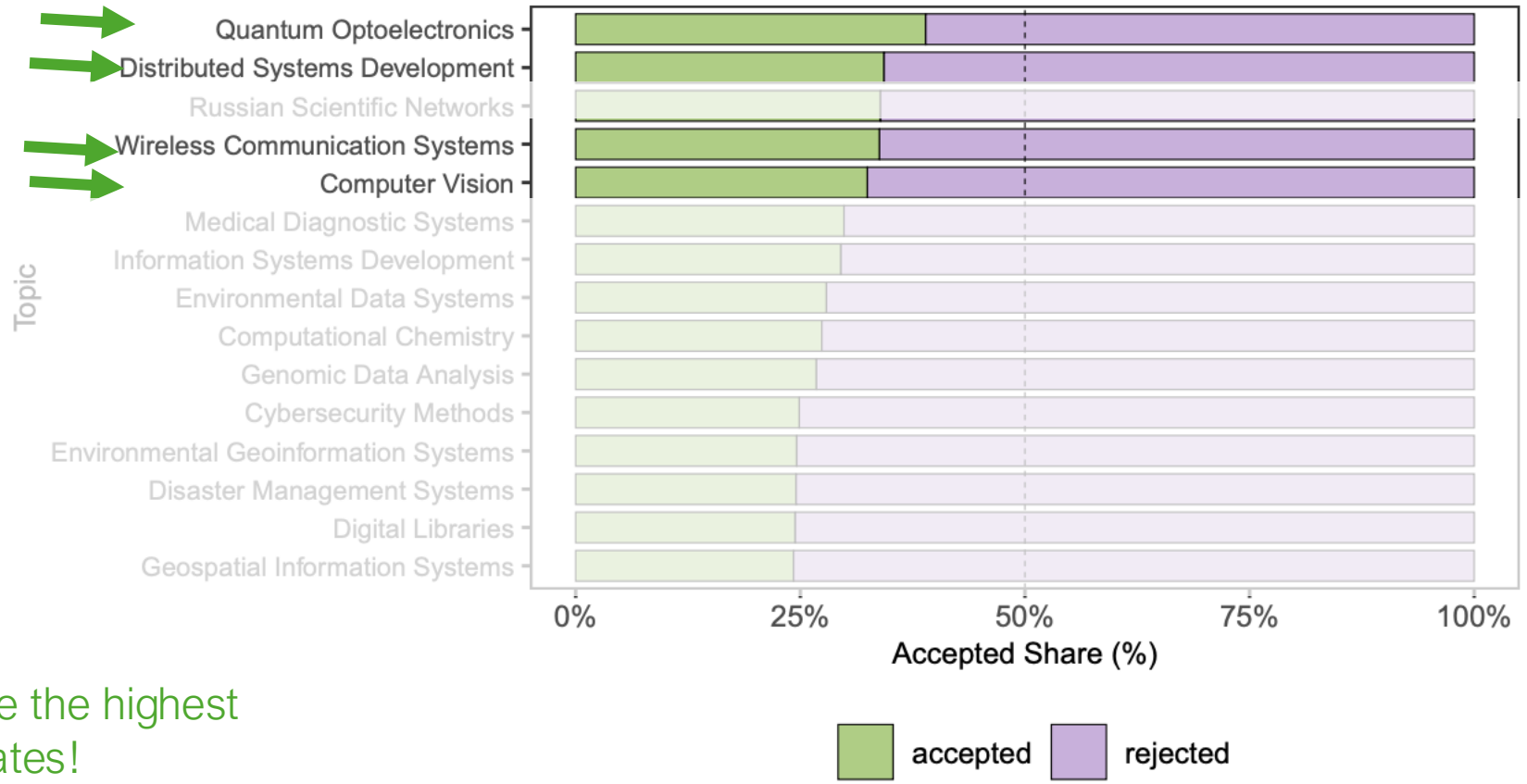


The most male-dominated topics have increased over time

Share accepted by topic

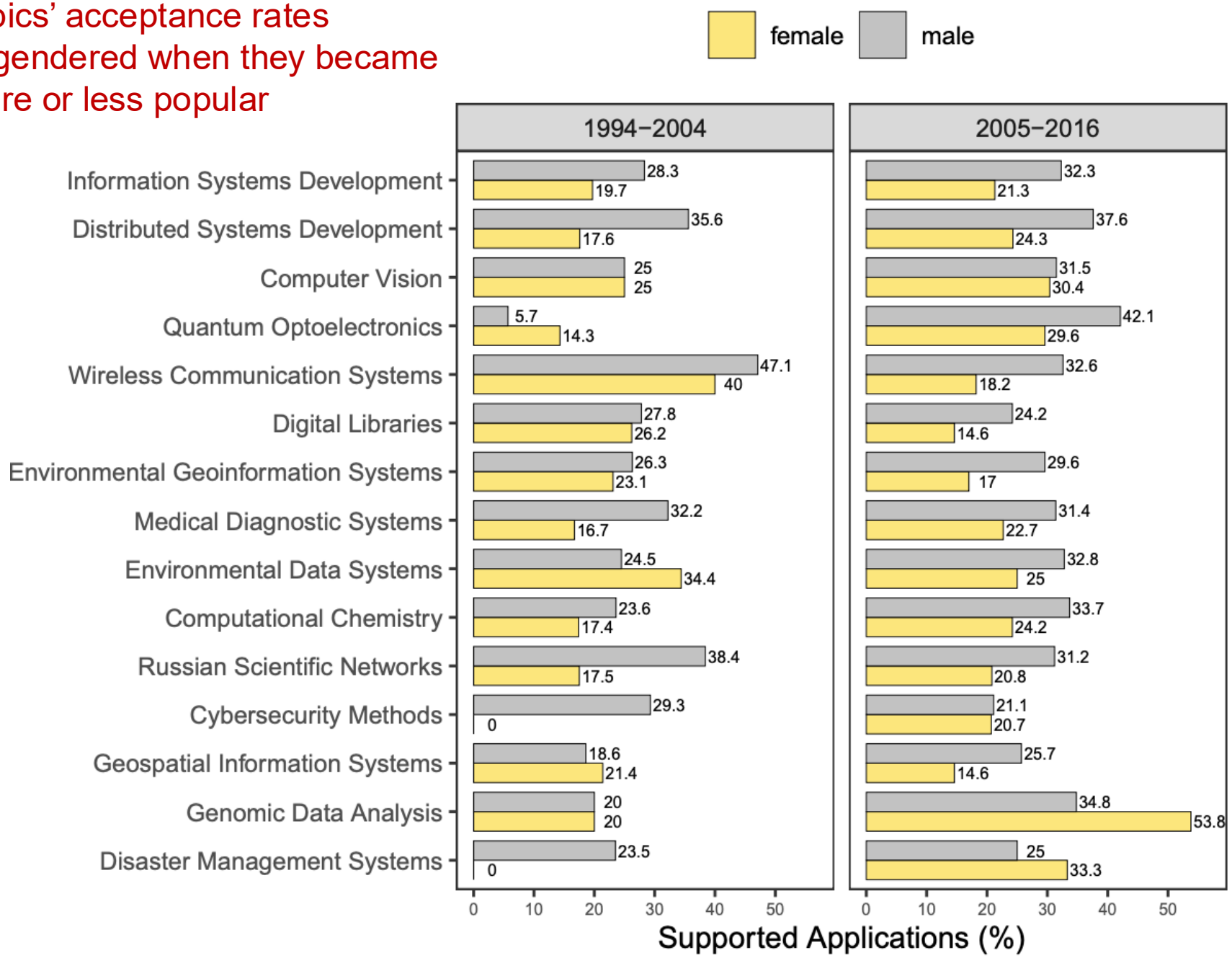


Share accepted by topic

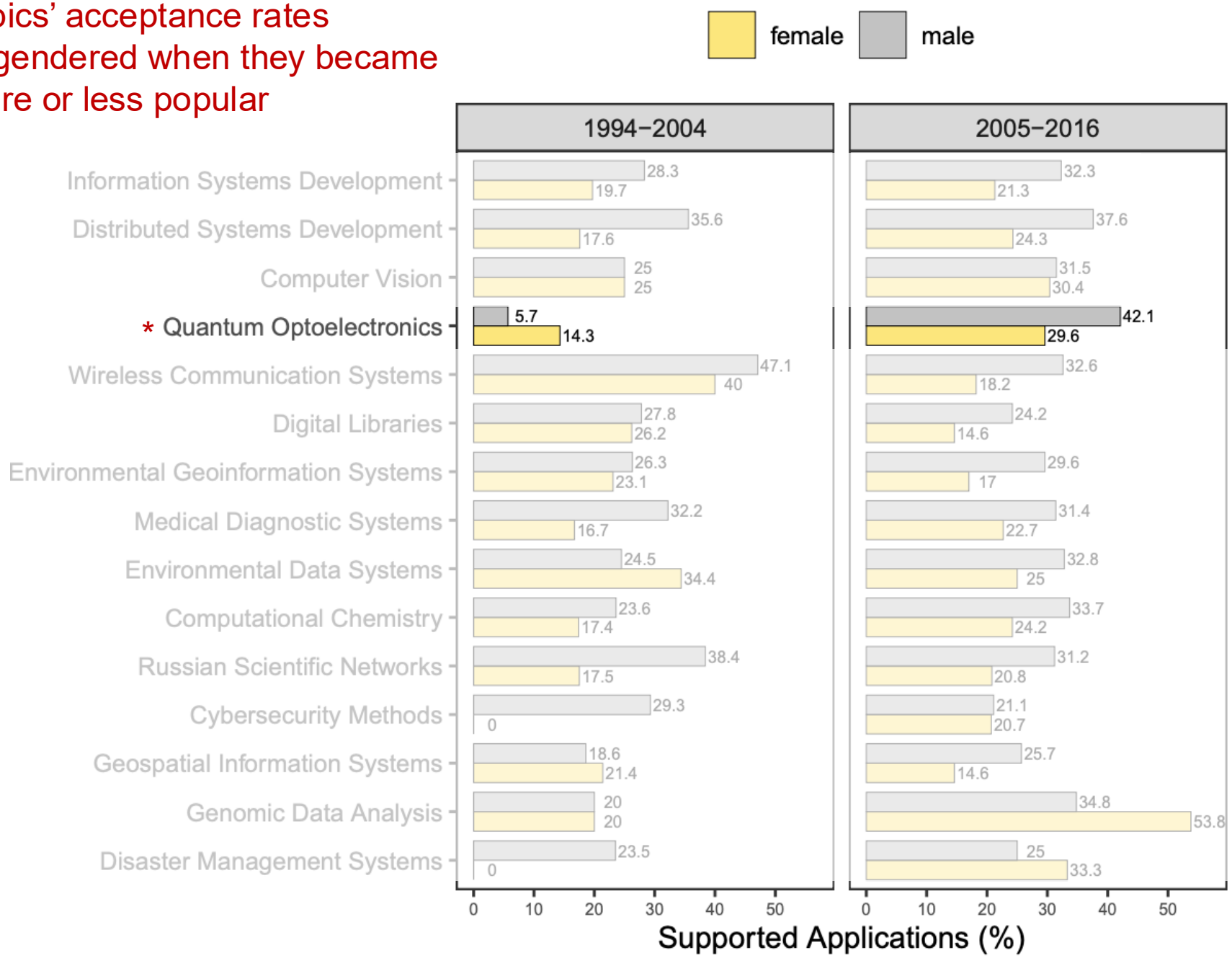


They also have the highest acceptance rates!

Some topics' acceptance rates became gendered when they became much more or less popular

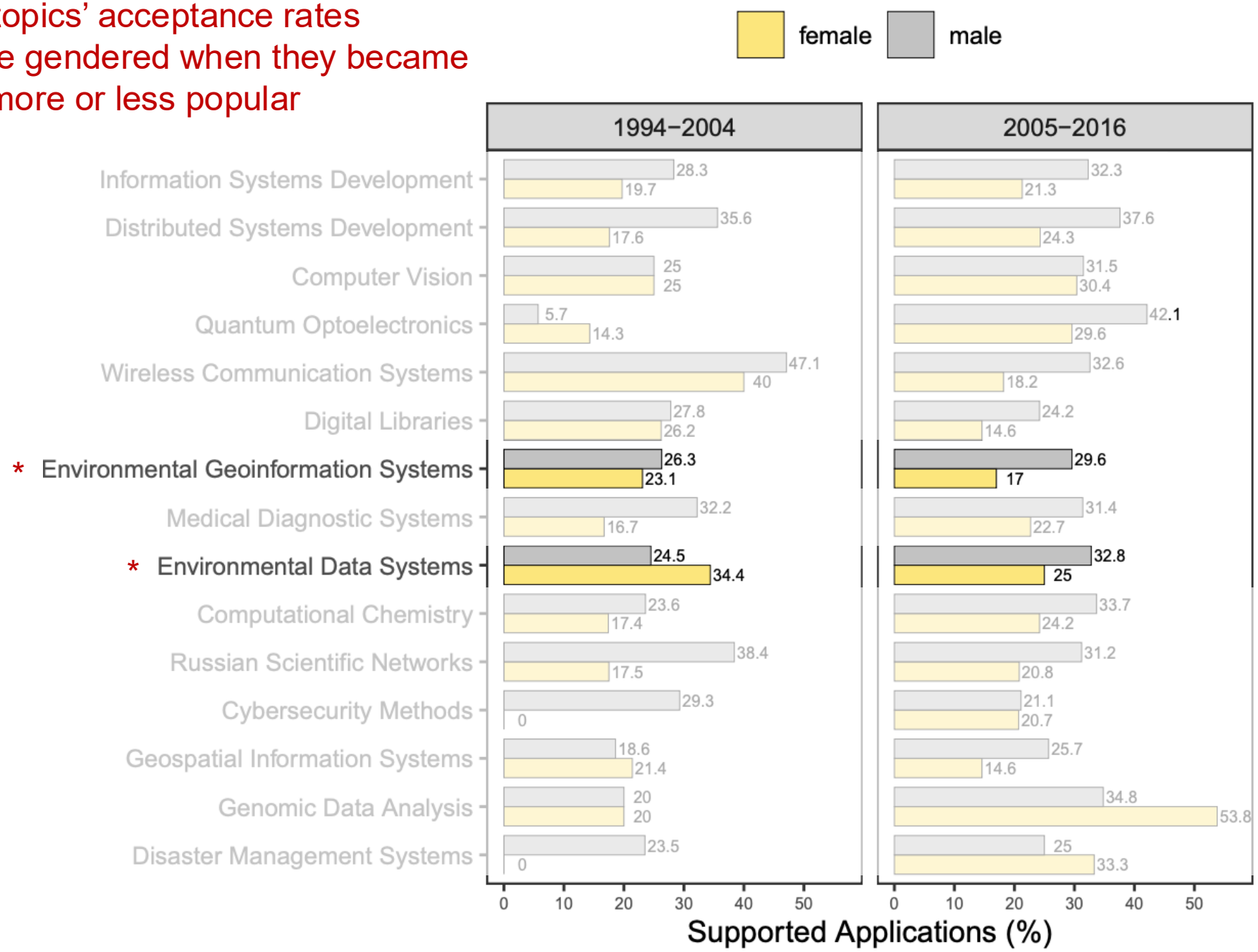


Some topics' acceptance rates became gendered when they became much more or less popular



Men were more likely to have proposals accepted from this topic

Some topics' acceptance rates became gendered when they became much more or less popular



Acceptance rates were equal or in favor of women, but became in favor of men when funding became less sparse

Takeaways

- Are changes in topics visible in the longitudinal data and do they correspond to sociopolitical periods?
  - Priority changes in funding are visible in longitudinal data – 2005 was an important economic and political year in Russia

Takeaways

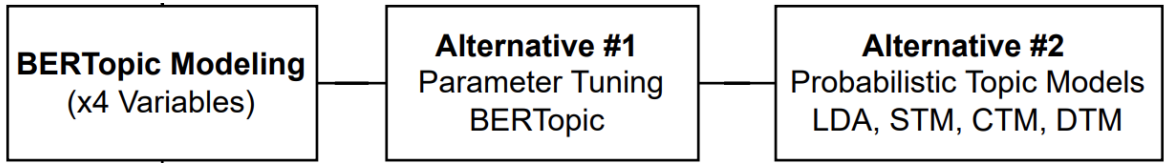
- Are changes in topics visible in the longitudinal data and do they correspond to sociopolitical periods?
  - Priority changes in funding are visible in longitudinal data – 2005 was an important economic and political year in Russia
- Are under-funded topics those that women disproportionately study?
  - Some topics are gendered because they have fewer apps (but acceptance rates are equal)
  - Some topics have gendered acceptance rates

Takeaways

- Are changes in topics visible in the longitudinal data and do they correspond to sociopolitical periods?
  - Priority changes in funding are visible in longitudinal data – 2005 was an important economic and political year in Russia
- Are under-funded topics those that women disproportionately study?
  - Some topics are gendered because they have fewer apps (but acceptance rates are equal)
  - Some topics have gendered acceptance rates
- In particular, the topics that became prioritized were already gendered. A few became more gendered. Untangling the drivers of these patterns is future work
  - Very important in today's funding climate in the US

Next steps

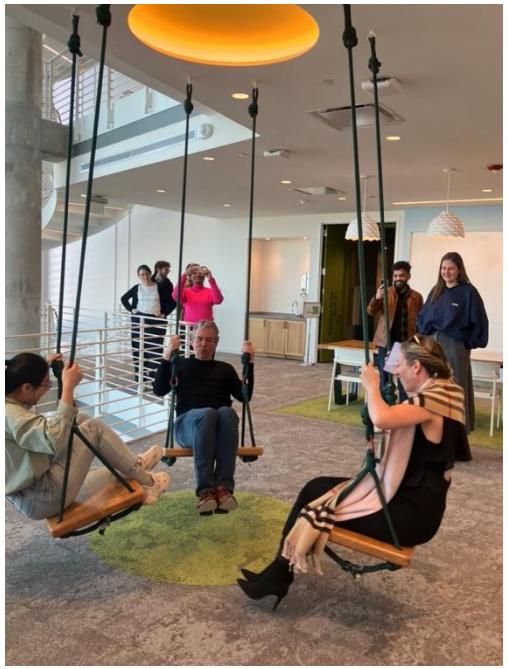
- Compare topic modeling across the four models (title en, title ru, title + abstract en, title + abstract ru)
- Is BERTopic the right strategy to use?
- Validate with “title of competition” (funding call)
- Compare to Open Alex (fit papers on our topics)
  - What topics are published but not funded?
- Scale topic modeling for all fields
- Predict whether or not a proposal is funded given gender, early career, field, year, etc.
- Compare to other countries



Thank you

Atlanta Academy on Science and Innovation Policy

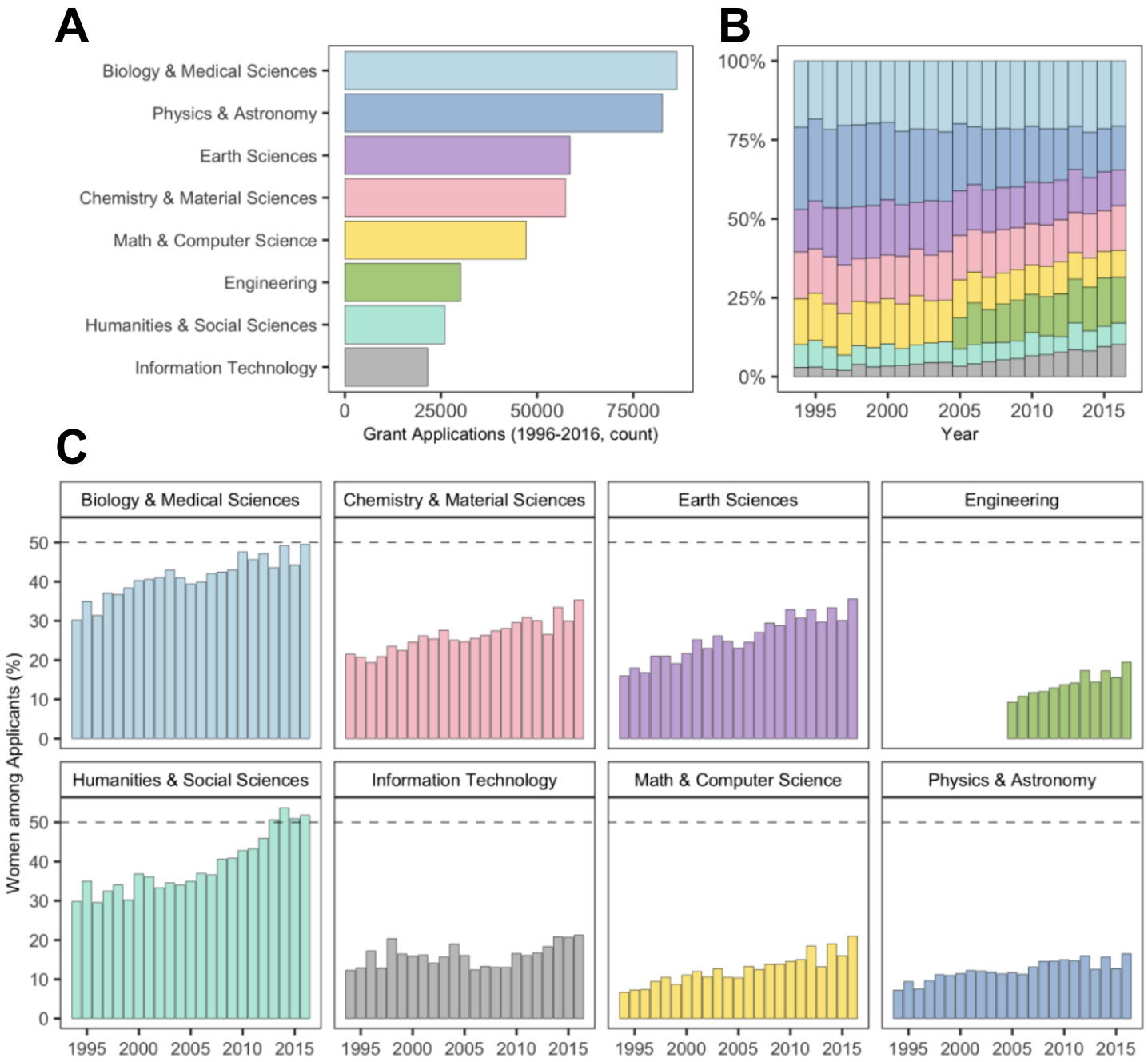
Xinzhe Li Vinicius Muraro

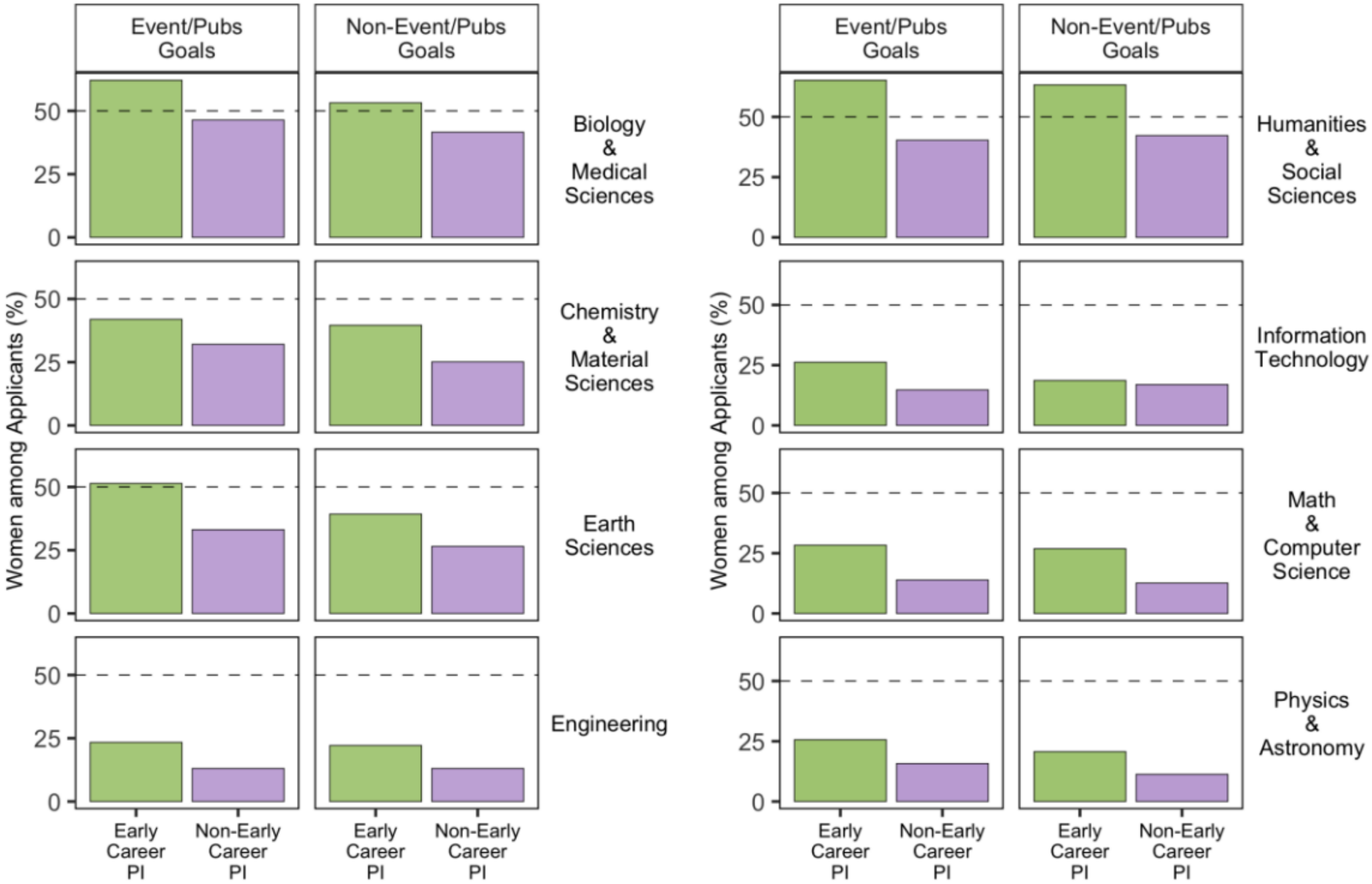


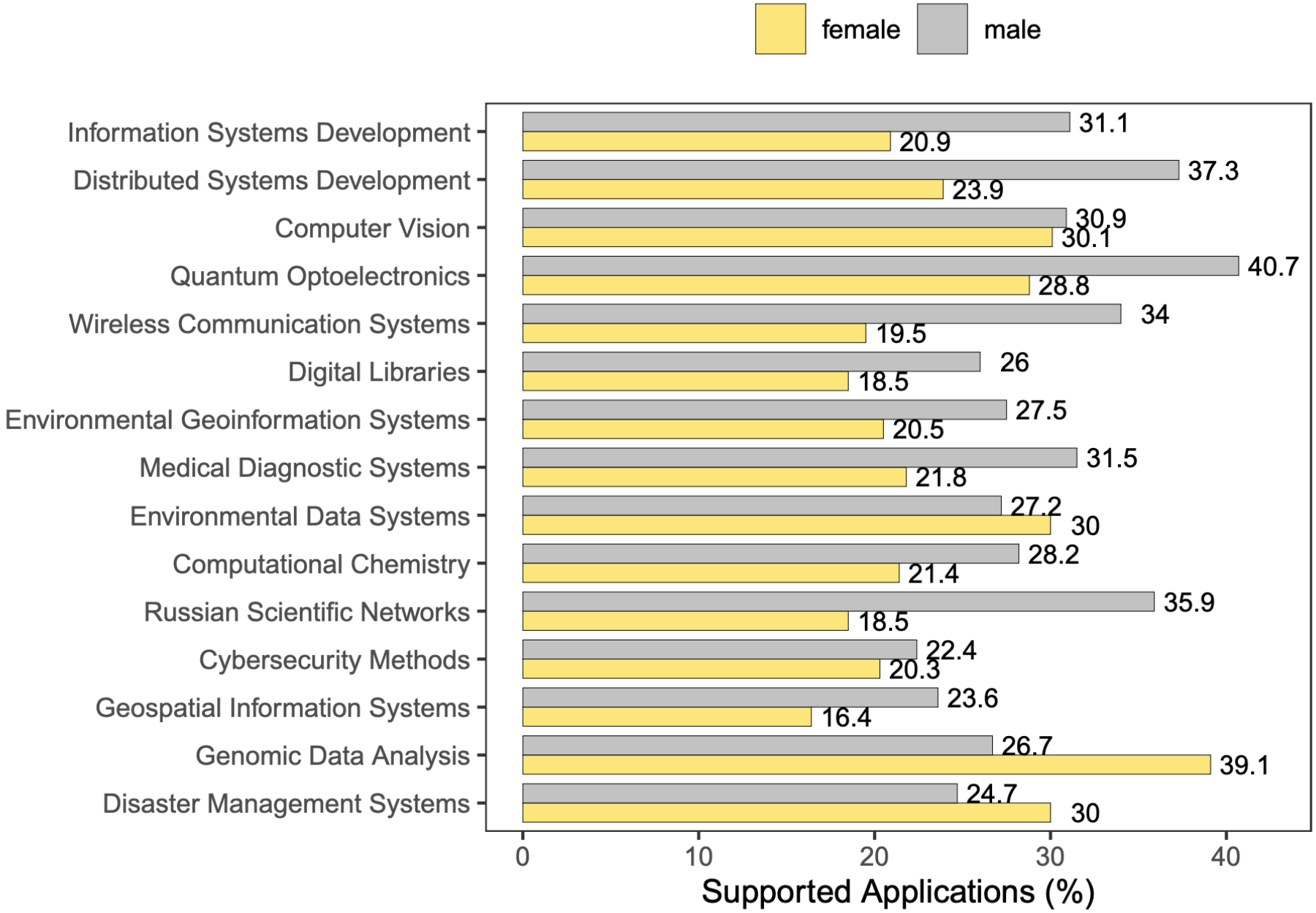
Jimmy and Rosalynn Carter School of Public Policy

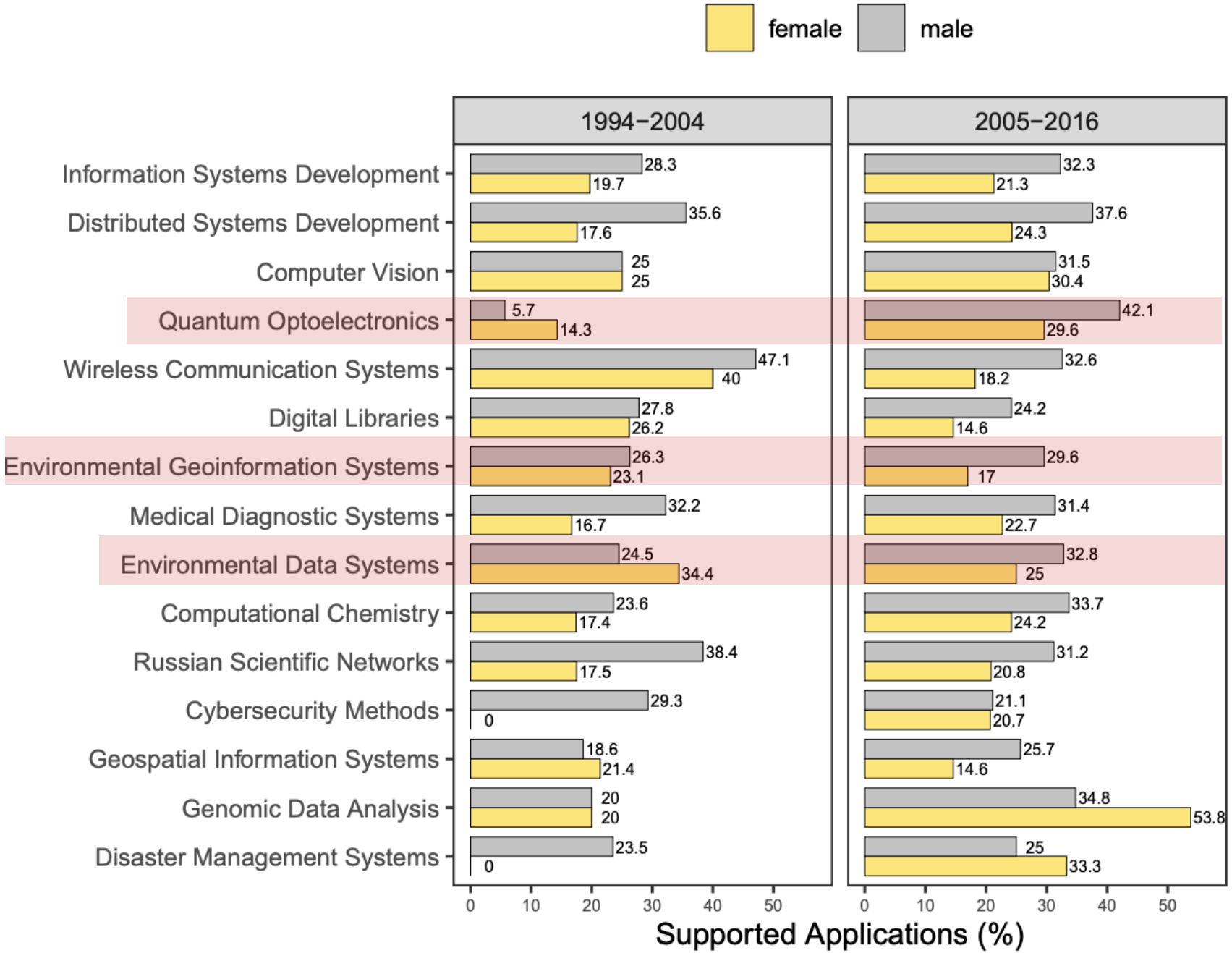
Cassidy Sugimoto

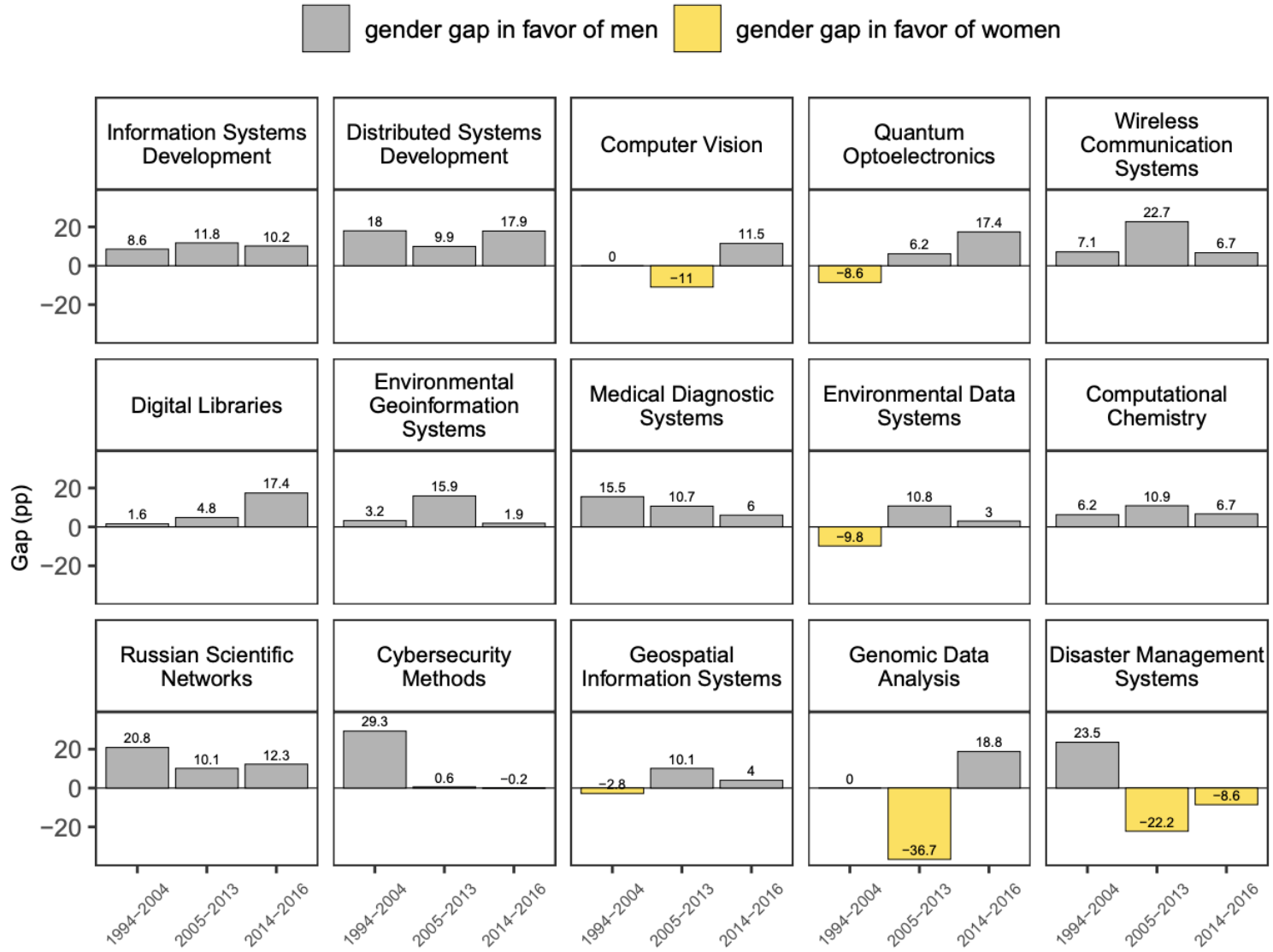
katiespoon.github.io/
kspoon.bsky.social
katherine.spoon@colorado.edu


A
Biology & Medical Sciences
Physics & Astronomy
Earth Sciences
Chemistry & Material Sciences
Math & Computer Science
Engineering
Humanities & Social Sciences
Information Technology
0
25000
50000
75000
Grant Applications (1996-2016, count)
B
100%
75%
50%
25%
0%
1995
2000
2005
2010
2015
Year
C
Biology & Medical Sciences
Chemistry & Material Sciences
Earth Sciences
Engineering
Humanities & Social Sciences
Information Technology
Math & Computer Science
Physics & Astronomy
Women among Applicants (%)
0
10
20
30
40
50
1995
2000
2005
2010
2015


Event/Pubs Goals
Non-Event/Pubs Goals
Women among Applicants (%)
50
25
0
Biology & Medical Sciences
Chemistry & Material Sciences
Earth Sciences
Engineering
Humanities & Social Sciences
Information Technology
Math & Computer Science
Physics & Astronomy
Early Career PI
Non-Early Career PI


female
male
Information Systems Development
31.1
20.9
Distributed Systems Development
37.3
23.9
Computer Vision
30.9
30.1
Quantum Optoelectronics
40.7
28.8
Wireless Communication Systems
34
19.5
Digital Libraries
26
18.5
Environmental Geoinformation Systems
27.5
20.5
Medical Diagnostic Systems
31.5
21.8
Environmental Data Systems
27.2
30
Computational Chemistry
28.2
21.4
Russian Scientific Networks
35.9
18.5
Cybersecurity Methods
22.4
20.3
Geospatial Information Systems
23.6
16.4
Genomic Data Analysis
26.7
39.1
Disaster Management Systems
24.7
30
0
10
20
30
40
Supported Applications (%)


female
male
1994–2004
2005–2016
Information Systems Development
28.3
19.7
32.3
21.3
Distributed Systems Development
35.6
17.6
37.6
24.3
Computer Vision
25
25
31.5
30.4
Quantum Optoelectronics
5.7
14.3
42.1
29.6
Wireless Communication Systems
47.1
40
32.6
18.2
Digital Libraries
27.8
26.2
24.2
14.6
Environmental Geoinformation Systems
26.3
23.1
29.6
17
Medical Diagnostic Systems
32.2
16.7
31.4
22.7
Environmental Data Systems
24.5
34.4
32.8
25
Computational Chemistry
23.6
17.4
33.7
24.2
Russian Scientific Networks
38.4
17.5
31.2
20.8
Cybersecurity Methods
29.3
0
21.1
20.7
Geospatial Information Systems
18.6
21.4
25.7
14.6
Genomic Data Analysis
20
20
34.8
53.8
Disaster Management Systems
23.5
0
25
33.3
0
10
20
30
40
50
Supported Applications (%)



The idea behind Figure 7d is that it directly shows the difference between the two bars in Figure 7c—that is, the gap between the share of accepted applications from men and women. For example, for *Disaster Management Systems* in the period 2014–2016, Figure 7c shows 20% of accepted applications from men and 28.6% from women. Then, in Figure 7d, we see the result of $20 - 28.6 = -8.6$.