Count us in! Beating the Bias in the Data Science and AI Talent Pipeline.

Lynda Hardman

What do we know about female academic career progression in the Netherlands?

The Dutch Network of Women Professors (https://www.invh.nl) produces an annual monitor of career progression in the academic world in the Netherlands. This provides valuable data on the relative numbers of women and men at different stages of their academic careers at faculty level for Dutch universities. It also allows figures to be calculated on different aspects of career progression, for example, where the ‘glass ceiling is least transparent’ during the course of a full career. This data monitors the progression of women through their career. However, due to the imprecise categorisation of topics into faculties, i.e. umbrella terms like “Engineering” being used to encompass small fields such as Data Science and AI, the already small number of students and staff in these fields are further masked.

Furthermore, the process from graduate to senior member of faculty is an arduous one. It is even more so in fields such as Data Science and AI as there are so few students to start with, and they highly sensitive to fall-out that the probability of having a decent pipeline of female professors is very low.

What do we know about female Informatics students in 16 countries in Europe?

Informatics Europe (https://www.informatics-europe.org) publishes an annual report of statistics collected via national contact points. This gives only a partial view of the numbers of students and graduates in the countries for which this data is collected.

Due to this incomplete view, the current available data is only a representation of half of Europe and therefore not a fully viable way to analyse the proportion of females in this field. The data needs to be extended to cover more countries and also cover career stages after a student has graduated. This requires the collection and coordination of fine-grained data from all European countries. Given the complexity of specifying when a course is “Computer Science” or “AI”, Informatics Europe expends considerable effort to ensure a comparable representation of student numbers across countries through personal discussions with national representatives.

Why do we need to know about the career progression of women in Informatics in the Netherlands and across Europe?

While Data Science and AI are fundamental to our digital society there are currently too few Data Scientists and AI practitioners generally and women are severely under-represented. This has consequences for women’s careers, for research topics and for societal applications of these technologies.

While many European countries collect data on students, this is not sufficiently fine-grained to determine how many women across Europe study, graduate and work in Data Science, AI and other Computer Science sub-fields. Our call is for organisations, including governments, across Europe to collect sufficiently fine-grained education and graduation statistics for AI, Data Science and other Computer Science sub-fields and make these widely available. Similar data is needed for other fields where there is systematic under-representation of women, or indeed men.

This data can then be used to monitor the effectiveness of measures that need to be taken to ensure that:
- the number and proportion of girls studying these fields increases;
- teachers at all school levels are able to teach these fields;
- educators at tertiary level are aware of the potential (unconscious) biases that may discourage women from studying or working in these fields;
- employers (academic and commercial) are aware of the needs of minority-friendly workplaces to ensure women remain working in these fields.

Given the current and future importance of Data Science and AI in our economy and society we cannot wait to tackle these problems at uncoordinated national levels.

International effort is required to collect and communicate best practices, and to monitor their effects with the goal of allowing women to take on a valuable and essential role in developing Data Science and AI applications and technologies and broader areas of Computer Science/Informatics.