Minerva Award 2024

Centre for Doctoral Training in Natural Language Processing, University of Edinburgh

10 Crichton Street, Edinburgh EH8 9AB, Scotland, United Kingdom

Contact: Dr Björn Ross, b.ross@ed.ac.uk, on behalf of the CDT Management Team

Abstract: The Centre for Doctoral Training (CDT) in Natural Language Processing (NLP) is a doctoral training programme for over 60 PhD students at the University of Edinburgh. The programme has successfully taken a range of measures to improve the recruitment and retention of female students. As a result 47% of students accepted into the programme were women or non-binary – much higher than the UK-wide average for a programme in this subject area – and students have expressed a high degree of satisfaction with the programme and its commitment to diversity and inclusion.

Initiative

Background and environment

The Centre for Doctoral Training (CDT) in Natural Language Processing (NLP) at the University of Edinburgh is an interdisciplinary training programme for PhD students. It has recruited 68 PhD students, 30 of them women. All students receive fully funded scholarships, a grant to pay for research costs (such as training opportunities, travel or equipment), and a wealth of support from the programme. It sits in the School of Informatics, its primary affiliation, and the School of Philosophy, Psychology & Language Sciences.

The CDT programme is delivered by a **management team** consisting of multiple academic and professional services staff. At the helm of the management team is the Principal Investigator, **CDT President Prof. Mirella Lapata**, a world-leading researcher in the field of Natural Language Processing, and a woman. The members of the management team take on different roles, among which is the **dedicated CDT Champion for Equality, Diversity & Inclusion (ED&I)**, Dr Björn Ross. He has regular office hours for CDT PhD students and, along with other members of the management team, conducts regular update sessions to inform and engage CDT students on its activities, including ED&I activities. Each PhD student cohort elects representatives and these student representatives also meet regularly with members of the management team.

Recruitment

The CDT recruited between 11 and 16 PhD students each year in five distinct cohorts, from 2019 to 2023. Throughout this period we especially encouraged applications from women, who are historically under-represented in Informatics, using the following approaches:

- **Inclusive advertising:** We monitor the use of inclusive language in the PhD studentship advert. The gender coding tool we use reports no masculine words, whereas feminine-coded words are used in our adverts. The advert also includes an eligibility statement ("we particularly encourage applications from women, minority groups and members of other groups that are underrepresented in technology"). We also ensured that adverts included a link to a video showing the CDT President, Mirella Lapata (a woman), so that she could serve as a role model to potential women applicants. We also prominently placed a photograph on our web page that shows the first PhD cohort, with a 50:50 gender split. This resulted in a number of applications the following year from female applicants who were inspired by the gender balance portrayed in the photo.
- **Open Days and Recruitment Ambassadors:** We encouraged CDT students to volunteer to speak to candidates considering an offer to achieve high conversion rates. We believe it is important for students to have an opportunity to learn about the research and equality culture in the place they are applying to. Two of the three students recruited for this purpose were women. The opportunity to speak to existing students was specifically pointed out on the web page section about how to apply.
- **Targeted advertising:** We promoted the PhD funding opportunity through various networks and channels, including networks of under-represented groups.
- The **ED&I Champion** was **looped in** at each stage of the recruitment process to ensure that strong female candidates were not overlooked.
- Mandatory **Unconscious Bias Training** for all individuals involved in the PhD selection process (which the students are also required to take once they start).

Beyond recruitment

While ED&I in a postgraduate researcher/PhD training context is often thought of as a recruitment activity alone, for us it is a constant priority throughout.

When students first join, we **allocate a buddy** (an existing PhD student) of the same gender. The cohort induction also includes a student-organised pride picnic that we hope will help ensure students of all gender and sexual identities feel welcome in the CDT. As the picnic is one of the first events that new students attend, we believe it sets the tone and helps establish an inclusive culture from the start.

Students in the same cohort are **seated together** on adjacent desks in the same space of the Informatics building. Because we achieve a good gender balance in the CDT (more on that below), and have a higher proportion of female students than the School of Informatics as a whole, female CDT students work in an environment where they do not perceive themselves as a minority. The resulting group dynamic also promotes inclusive behaviour in male students.

CDT students **regularly meet** with the management team including the ED&I Champion (in yearly update sessions for all students and two student representative meetings per semester). At each of these meetings, we **re-emphasise to the students our commitment** to making the CDT an inclusive place and we encourage students to promptly share any issues they may be facing. These issues and their possible solutions are then discussed at the meetings, and a management team member is tasked with following up and solving each specific issue. To give an example of such an issue, free period products were made available on all floors of the Informatics building after their limited availability was raised by a CDT student.

Our **overall strategy** overall has been to empower students (in a bottom-up approach) instead of imposing events on students (top-down). We make clear to students that we will support them with any ED&I related concerns, and that resources exist to support their ideas.

This has led to several events and initiatives: (1) a student-led **ED&I bookclub** for CDT students (funded by the CDT). An example of a book the students have chosen is "How Men Can Help: A Guide to Undoing Harm and Being a Better Ally" by Sophie Gallagher. (2) a student-led **ED&I reading group** for both students and staff where ED&I issues in research are discussed (funded by the School). This group was started by CDT in NLP students and it is open to and attended by members of the entire School of Informatics, showing how our philosophy is impacting the Informatics community beyond the CDT. (3) At the recent, student-led **CDT conference**[1] (jointly with the CDT in Speech and Language Technologies at the University of Sheffield) two of three keynote speakers were female, and the only non-female speaker focused on an ED&I subject.

Other activities. The CDT and the School of Informatics encourage female students to engage with networking events and we encourage the sharing of good practice by students/staff by attending events and joining networks, e.g., Women in Machine Learning, ACMW network, Black in AI, Women in NLP. For example, multiple CDT students have been organisers, panellists or participants at the Hoppers International Women's Day Event held at the School of Informatics, and part of the 2020 recruitment campaign included a Women in AI/University College London/CDT in NLP collaboration which CDT President Mirella Lapata hosted and was attended by two of our female student ambassadors. The CDT also runs a seminar series where students are encouraged to interact with speakers in student lunches, where 38% of recent speakers were women. (In comparison, only 23% of academic staff in Engineering & Technology across the UK are women[2].)

---

[1] https://cdtjointconference.notion.site/Keynote-Speakers-091f35a145d0476e9a0d132a9ab26760
[2] https://www.hesa.ac.uk/data-and-analysis/staff/areas

Evidence

Recruitment: A high proportion of female students

Year	Women	Men	Non-binary or not disclosed	Total
2019	8 (50%)	8	0	16
2020	5 (36%)	8	1	14
2021	5 (36%)	8	1	14
2022	5 (38%)	8	0	13
2023	7 (64%)	4	0	11
Total	30 (44%)	36	2	68

Table 1. Number of students accepted into the CDT in NLP programme by gender and cohort

Female students make up 44% and female and non-binary students together make up 47% of the PhD students in the programme.

For comparison, the most recent HESA data[3], from 2021/22, indicates that in the Computing field in the UK, only 29.5% of postgraduate research students (i.e., students studying towards a PhD and similar degrees) are women. For Scotland, the number is 27.9%.
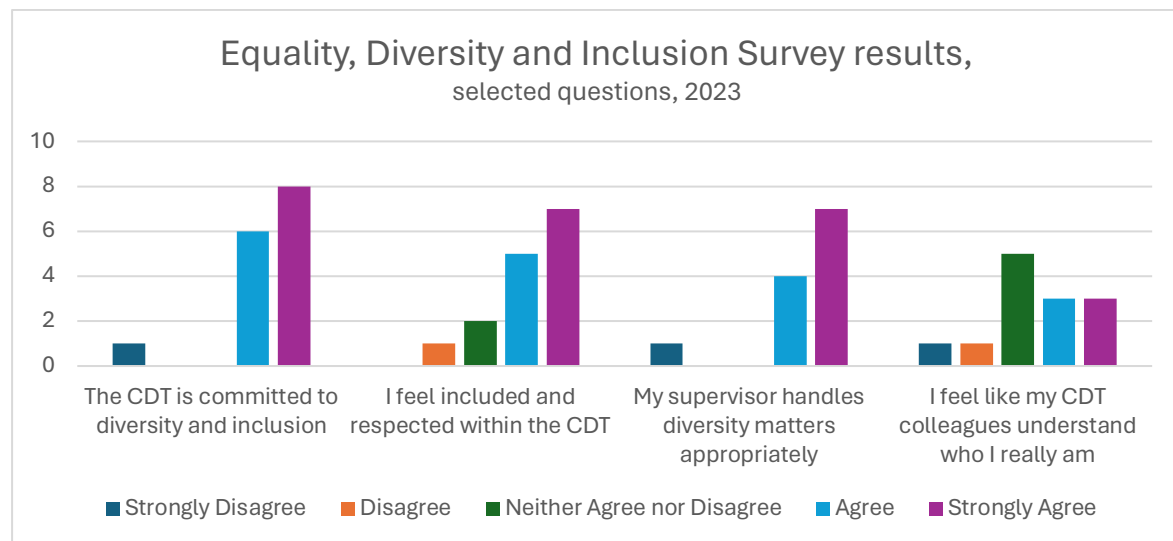
Year	Women	Men	Non-binary or not disclosed
2019	89%	67%	n/a
2020	83%	80%	100%
2021	83%	62%	100%
2022	63%	67%	n/a
2023	86%	56%	n/a
Total	81%	66%	100%

Table 2. Conversion rates of students for the CDT in NLP programme by gender and cohort (no. of offers made / no. of offers accepted)

Our analysis of the application numbers shows that high number of female students is partly due to an excellent conversion rate for female students: over the four years analysed, 81% of female offer holders accepted the offer of a PhD place in the CDT in NLP, in comparison with 66% of male offer holders. We believe that this shows how the CDT in NLP has succeeded in presenting a welcoming environment to female applicants.

---

[3] https://www.hesa.ac.uk/data-and-analysis/sb265/figure-13

Beyond recruitment: Student satisfaction with diversity and inclusion matters

Equality, Diversity and Inclusion Survey results,
selected questions, 2023

In 2021 and 2023 we conducted Equality, Diversity and Inclusion Surveys among our CDT students to assess our progress. On a five-point Likert scale between "Strongly disagree" and "strongly agree", by 2023, 93% of respondents agreed or strongly agreed that "The CDT is committed to diversity and inclusion" (up from 80% in 2021). On the question, "I feel included and respected within the CDT", by 2023 80% of students agreed or strongly agreed (up from 67% in 2021). The overwhelming majority of students expressed not only satisfaction with CDT management but also with their PhD supervisors with regards to ED&I topics. On the question, "My supervisor handles diversity matters appropriately", 92% agreed or strongly agreed in 2023 (up from 85% in 2021).

The survey results also helped us identify areas of improvement. For example, in 2021, while the COVID pandemic was in full swing, 27% of students disagreed or strongly disagreed with the statement that "I feel like my CDT colleagues understand who I really am." As a result, we prioritised cohort-building activities, such as a yearly joint trip to an outdoor activity centre. In 2024 this was followed by a student-led CDT writing retreat (fully funded by the CDT) where three of the four organisers were women. By 2023, the number of respondents disagreeing or disagreeing strongly with the statement had reduced to 15%.