Gendered Words and Grant Rates: A Textual Analysis of Disparate Outcomes in the Patent System

Deborah Gerhardt, Miriam Marcowitz-Bitton, W. Michael Schuster, Avshalom Elmalech, Omri Suissa & Moshe Mash[1]

Text is a vehicle to convey information that reflects the writer's linguistic style and communicative patterns. By studying these attributes, we can discover latent insights about the author and their underlying message. This article uses such an approach to better understand patent applications and their inventors.

While prior research focuses on patent metadata (i.e., filing year or gendered inventor names), we employ machine learning and natural language processing to extract hidden information from the words in patent applications. Through these methods, we find that inventor gender can often be identified from textual attributes—even without knowing the inventor's name. This ability to discern gender through text suggests that anonymized patent examination—often proposed as a solution to mitigate disparities in patent grant rates—may not fully address gendered outcomes in securing a patent.

Our study also investigates whether objective features of a patent application can predict if it will be granted. Using a classifier algorithm, we correctly predicted whether a patent was granted 60% of the time. Further analysis emphasized that writing style—like vocabulary and sentence complexity—disproportionately influenced grant predictions relative to other attributes such as inventor gender and subject matter keywords.

Lastly, we examine whether women disproportionately invent in technological areas with higher rejection rates. Using a clustering algorithm, applications were allocated into groups with related subject matter. We found that 85% of female-dominated clusters (over 50% women inventors) have abnormally high rejection rates, compared to only 45% for male-dominated groupings.

These findings highlight complex interactions between textual choices, gender, and success in securing a patent. They also raise questions about whether current proposals (e.g., anonymized examination) will be sufficient to achieve gender equity and efficiency in the patent system.

Keywords: Patent Law, Intellectual Property, Gender Disparities, Textual Analysis, Natural Language Processing (NLP), Machine Learning, Gender Identification

---

[1] Deborah R. Gerhardt is the Paul B. Eaton Distinguished Professor of Law at the University of North Carolina School of Law.
Miriam Marcowitz Bitton is the Raoul Wallenberg Human Rights Full Professor of Law.
Mike Schuster is an associate professor at the University of Georgia's Terry College of Business with a courtesy appointment at the University of Georgia School of Law.
Avshalom Elmalech is an associate professor of Information Science at Bar Ilan University.
Omri Suissa holds a PhD in artificial intelligence and natural language processing from Bar Ilan University. He is currently a research scientist at Brown University and a co-founder and VP of R&D at ClearMash Solutions.
Moshe Mash holds a PhD in artificial intelligence from Ben-Gurion University. He is the founder of MashInnovateAI.