

Trust and data ownership

Can online content indicate an individual's 'real-life' personality?

Research results

PLAIN TEXT PREVIEW VERSION

Alice Matthews B.Comm/B.A University of Sydney, GradDip Analytics UNSW,
Data analyst and statistician

Andrew Hine B.Sc/M.Sc. University of Oxford, Founder and CEO Reputonaire &
TechieRank

About the authors

Alice Matthews B.Comm/B.A. History and Economics (University of Sydney), Grad.Dip.Analytics (UNSW) is a data scientist currently undertaking a master's degree in data science at UNSW.

Andrew Hine MEng Engineering and Computer Science (University of Oxford) is co-founder and CEO of Reputacionaire, an award-winning RMIT start-up that strives to improve people's lives by empowering them to own and leverage their online data, while working towards 7 of the UN's 17 sustainability goals. Andrew's Masters' degree at Oxford was sponsored by the UK Intelligence Service, where he later worked, sparking his interest in data ownership and security.

Executive Summary

The global market today is characterised by distance and separation, making it harder to connect meaningfully with clients, talent and colleagues. In addition, the rise in global migration flows over the past decades has resulted in large populations of people who are dislocated from their home networks and community, and must find opportunities in their host country without local references or experience.

Inferring personality from online content offers an opportunity to bridge the gap that separates people, and will become a fundamental part of the way we interact in the years to come.

To support this vision, Reputacionaire has conducted a data-driven investigation into the correlations between online content-based personality inference and psychometric survey personality measurements, in order to determine **whether online social media content can be a good approximator of more traditional psychometric surveys**. Our results indicate that text content from a range of online sources can indeed reliably estimate the results of the IPIP 50-item psychometric survey.

We used [IBM's Personality Insights](#) service (discontinued as of December 2021) and the International Personality Item Pool (IPIP) 50-item [Big 5/OCEAN model survey](#), and found a statistically significant ($p < 0.05$) **positive correlations of 0.34 (34%) in Conscientiousness** scores between the two methods, and **0.36 (36%) in Extraversion** scores. Both these traits [have been linked](#) to employability. We also found strongly predictive ordinary least squares regression models linking online-content inferred personality analysis to psychometric survey personality analysis.

This finding illustrates that 'real life' personality traits can reasonably be inferred from an individual's online content. **This supports Reputacionaire's vision that an individual's online presence is an asset that should be available for them to own and leverage in support of their real life goals.** As such, our findings have implications in the fields of recruitment and employability, and may save time and money for recruiters while increasing employability for applicants.

This paper will detail the methodology and findings of our research, and offer some comments on the importance of these findings in today's market. We will further outline how **Reputacionaire enables assessors to provide their applicants the option of plugging in verified insights from their online profiles, saving time and making it easier for them to be more inclusive.**

Key Points

- ❖ Many individuals nowadays have a large body of published content in their public online profiles. We believe that this investment in online spaces should have a 'real world' payoff for individuals.
- ❖ NLP prediction of personality traits generated from an individual's online public content with an IPIP 50-item psychometric survey assessment for 101 individuals.
- ❖ Data analysis methods found **positive correlations between our automatic online content-based measure of personality traits and the standard IPIP 50-item manual survey method** in Conscientiousness, Extraversion and Openness.
- ❖ Ordinary Least Squares regression analysis showed online content-based personality scores could strongly predict scores from the psychometric survey, even when generalised to text types other than the tool's native Twitter.
- ❖ Our results indicate robustly that **analysing online content may be a reasonable proxy for traditional survey methods to reveal Big 5 personality traits.**
- ❖ Inferring personality from online content offers an opportunity to bridge the gap that separates people, and will become an even more fundamental part of the way we interact in the years to come.
- ❖ Assessing soft skills is vital in finding the right talent, but we lack the means to do so quickly, easily and accurately. Online content analysis may offer a solution.
- ❖ Online content analysis may offer a means of assessing candidates that is blind and impartial, and therefore unbiased.
- ❖ Privacy is a key consideration when it comes to social media insights.
- ❖ Individuals who move first to leverage this opportunity by claiming ownership of their online reputation, and organisations that allow individuals to submit it as part of their assessments will gain first-mover advantages, saving time and money on the candidate assessment process. Reputationaire provides tools for assessors to provide their applicants the option of plugging in verified insights from their online profiles saving assessors time and making it easier to be more inclusive.

Want to access the full research?

Everyday our individual lives become more digital and online.

How can you interpret other's online profiles and behaviour and how others perceive you from your online content?

And for organisations the "cost of maintaining trust consumes 35% of the US workforce - and in financial/professional services it consumes 48%".

How can including novel online data sources reduce your cost of doing business and lead to better and more diverse selections?

Head to reputationaire.com/research to access the full 47 page graphical version, with the complete analysis, use case examples, full model data, supporting research and more.