

Traços de personalidade e Big Data: desafios éticos a partir de pesquisas realizadas com a base *myPersonality*. Num cenário em que dados digitais transformam a maneira como compreendemos o comportamento humano, este estudo analisou uma iniciativa pioneira, o projeto *myPersonality*, para investigar como e para que fins a avaliação da personalidade é realizada na era do Big Data. O aplicativo *myPersonality*, utilizando a escala IPIP, desenvolveu uma base de dados com informações de mais de seis milhões de usuários do Facebook. Este artigo analisou 23 pesquisas realizadas a partir dessa base. Foram identificados 13 alvos de investigação, como personalidade e linguagem; analisando 73 atributos disponíveis em redes sociais; e os estudos utilizaram 52 ferramentas metodológicas, desde modelos estatísticos a algoritmos avançados. A pesquisa concluiu que a avaliação da personalidade foi automatizada visando aplicações relacionadas com a adaptação de produtos e serviços às características individuais e que o uso intensivo de dados digitais suscita questões éticas que obrigam a uma reflexão sobre as responsabilidades científicas e sociais no manejo desses dados.

PALAVRAS-CHAVE: personalidade; Big Five/Cinco Grandes fatores; análise de Big Data; psicologia; ética em pesquisa.

Personality traits and Big Data: ethical challenges arising from research conducted using the *myPersonality* dataset.

In a world where digital data is reshaping our understanding of human behavior, this study examined a pioneering initiative, the *myPersonality* project, to explore how personality assessment is carried out in the Big Data era. The Facebook *myPersonality* app, which employed the IPIP scale, built a database containing information on more than six million Facebook users. This article analyzed 23 *myPersonality*-based publications to understand how personality is studied in the Big Data context. Researchers identified 13 lines of inquiry, such as personality itself and language, drawing on 73 social-media attributes and employing 52 methodological tools, from traditional statistical models to advanced algorithms. The study concludes that personality assessment has been automated to tailor products and services to individual characteristics, and that the intensive use of digital data raises ethical concerns, demanding reflection on the scientific and social responsibilities involved in handling such data.

KEYWORDS: personality; Big Five; Big Data analysis; psychology; ethics in research.

THARIK REIS

CAMILA BRUNING

KARLO MARQUES JUNIOR

Traços de personalidade e Big Data: desafios éticos a partir de pesquisas realizadas com a base *myPersonality*

INTRODUÇÃO

A análise dos processos psicossociais associados aos fatores por meio dos quais as diferenças individuais se manifestam naquilo que se denomina “personalidade” desempenhou um papel fundamental no desenvolvimento e reconhecimento da psicologia como ciência e profissão. Ao longo dos anos, os avanços científicos e as controvérsias acerca das práticas psicométricas estiveram em consonância com o contexto histórico e social nos quais ocorreram. É de se esperar que o mesmo aconteça numa época em que a produção de dados sobre diferentes aspectos da vida social – e a tecnologia para os analisar – se desenvolve num ritmo acelerado.

Neste contexto, o presente trabalho teve como objetivo compreender como, e para que fins, os fatores sociais nos quais a personalidade se manifesta são estudados no contexto de *Big Data*. Optou-se por abordar essa questão a partir da análise de um caso específico: o das pesquisas que analisaram dados da base chamada *myPersonality*.

As perguntas que guiaram a investigação foram, portanto: (i) a partir de que procedimentos metodológicos os fatores sociais nos quais a personalidade se manifesta foram estudados nas pesquisas que analisaram dados da base *myPersonality* e (ii) quais as finalidades declaradas desses estudos.

A evolução histórica da psicologia, sobretudo na área da psicometria e do estudo da personalidade, já demonstrou o poder de métodos empíricos para mensurar traços individuais por meio de escalas e questionários (Swan et al., 2023; Stefana et al., 2025). Com o advento do Big Data e o desenvolvimento

de novos métodos de análise, há oportunidades para que esse paradigma se expanda, ao integrar técnicas computacionais avançadas na recolha e análise de informações digitais em grande escala.

Processos clássicos de avaliação da personalidade e do seu impacto na vida social, que se desenvolveram ao longo do século xx, podem agora ser complementados e potencializados pela capacidade dos algoritmos de processar volumes massivos de dados extraídos de redes sociais, permitindo uma compreensão dinâmica e em tempo real dos comportamentos humanos e sociais (Habib et al., 2024). Decorre disso uma intersecção entre os temas psicometria, características de personalidade, Big Data e suas implicações para a análise dos fenómenos sociais, evidenciando-se uma continuidade histórica entre os métodos de pesquisa em psicologia e nas ciências sociais e a necessidade contemporânea de repensar as implicações sociais, éticas e metodológicas neste novo cenário, no qual o acesso aos dados pessoais ocorre numa escala sem precedentes.

CONTEXTUALIZANDO O ESTUDO DA PERSONALIDADE NA PSICOLOGIA

O estudo sistemático dos fatores de personalidade e do seu subsequente impacto na compreensão dos fenómenos sociais foi inaugurado no início do século xx e consolidou-se com a obra de Allport (1937), que definia a personalidade como “the dynamic organization within the individual of those psychophysical systems that determine his unique adjustments to his environment” (Allport, 1937, p. 48). A chamada “teoria dos traços”, proposta originalmente por Allport, propõe que a personalidade é composta por um conjunto de traços relativamente estáveis, identificáveis a partir da análise léxica dos termos descritivos da linguagem (Allport e Odbert, 1936), além de identificar termos descritivos da personalidade, levando ao desenvolvimento do modelo dos cinco grandes fatores (Passos, 2014), cuja hipótese tem sido corroborada por meio de evidências empíricas consistentes em diversos trabalhos desde então (McAdams, 1997; Lambiotte e Kosinski, 2014). Allport e Odbert (1936) iniciam essa tradição ao identificar e catalogar uma vasta lista de traços de personalidade a partir da análise léxica da linguagem. A abordagem pioneira estabeleceu as bases para o entendimento de que a compreensão da personalidade se poderia dar por via do desmembramento em componentes distintos extraídos da maneira como as pessoas se descrevem e descrevem os outros.

Posteriormente, Cattell (1943), influenciado por essa fundação, avançou ao aplicar técnicas estatísticas e de análise fatorial para reduzir a lista de traços a um número gerível e teoricamente significativo. O autor justificou a adoção

de 16 fatores com base numa argumentação empírica, utilizando técnicas de análise fatorial para explorar as correlações entre diversos descritores de personalidade. No seu trabalho argumentou que, ao identificar agrupamentos de traços que se correlacionavam de maneira consistente e reproduzível, era possível reduzir a complexidade dos dados sem perder a variabilidade essencial dos traços. Assim, os 16 fatores emergiram como dimensões primárias que capturam de forma não redundante a estrutura subjacente do constructo que se convencionou denominar “personalidade”, enquanto 8 fatores secundários foram identificados como refletindo uma hierarquia superior que integra esses agrupamentos, proporcionando uma estrutura sistematizada para explicar variações individuais no constructo.

Mesmo contando com um número significativamente menor de fatores, o modelo ainda era complexo e diversos autores se dedicaram a replicá-lo e a simplificá-lo. Outras investigações lexicais dos fatores aconteceram em diversos idiomas, como holandês, alemão, japonês, chinês, hebraico e russo (Hutz et al., 1998; John, Naumann e Soto, 2008), com pequenas discordâncias acerca do número de fatores (John, Naumann e Soto, 2008).

Os fatores não foram apenas traduzidos, mas adaptados especificamente a cada contexto. No caso da adaptação para a língua portuguesa consideraram-se os seguintes cinco traços: extroversão, socialização, realização, neuroticismo e abertura. Termos como “agreeableness” e “conscientiousness”, são empregados como “socialização” e “realização”, escolha que resulta de estudos que validaram essas nomenclaturas como mais adequadas para descrever as características comportamentais observadas em amostras brasileiras, promovendo uma maior validade dos instrumentos de avaliação nesse contexto específico (Hutz et al., 1998; Nunes e Hutz, 2002).

Os cinco fatores foram definidos da seguinte maneira (Nunes e Hutz, 2002): extroversão: o quanto e com que intensidade as interações sociais são preferidas e a estimulação é necessária; socialização: refere-se ao espectro, nas atividades de socialização, que abrange desde a compaixão até ao antagonismo; realização: o quanto uma pessoa seria capaz de se organizar e se manter firme na busca de um objetivo; neuroticismo: refere-se ao grau de instabilidade e desajuste emocional; abertura: refere-se ao grau de reconhecimento da importância de ter novas experiências e a frequência de comportamentos exploratórios.

A teoria dos traços de personalidade, sobretudo na sua formulação Big Five, tornou-se hegemónica na psicologia académica e aplicada, orientando desde pesquisas laboratoriais até processos de seleção profissional (Brüning, 2015). Contudo, essa primazia não está isenta de críticas, tendo sido já apontados dilemas éticos e políticos relevantes. Aguiar (2005) questiona a fixidez

atribuída aos traços, a validade, imparcialidade e representatividade dos testes, a sua aplicação em contextos artificiais, a falta de escrutínio das organizações sobre a qualidade dos instrumentos e o desinteresse potencial dos avaliados.

Mills (2015) vai mais além, apontando que esses modelos carecem de evidência objetiva de personalidade – existindo apenas nos próprios testes – e funcionam como construções teóricas imbricadas em interesses do capital, marcando uma mudança na forma de entender o “eu”. Segundo a autora, ao reduzir a complexidade humana a dimensões estáveis, a abordagem individualista omite o caráter social da personalidade, reforça o *status quo* das relações de poder e desloca o foco da transformação social para a autogestão dos indivíduos, revelando o potencial ideológico dessa abordagem ao individualizar problemas sociais, naturalizar hierarquias de poder e reduzir a subjetividade a uma mercadoria para o mercado de trabalho. Brüning (2015) aponta ainda que essa ênfase nos traços herdados ou aprendidos traduz uma visão mecanicista que pode naturalizar desigualdades e obscurecer fatores culturais e históricos na formação do comportamento. Essas objeções revelam que, apesar da sua presença hegemônica, mesmo na sua versão analógica, o modelo de traços já suscitava inquietações sobre justiça social, autonomia dos sujeitos e transparência metodológica, exigindo reflexões críticas sobre o seu uso em pesquisas e intervenções.

O ESTUDO DOS TRAÇOS DE PERSONALIDADE NO CONTEXTO DO BIG DATA E O PROJETO *MYPERSONALITY*

Vale a pena contextualizar o desenvolvimento dos instrumentos de pesquisa utilizados para a recolha de informações para a análise dos cinco fatores de personalidade. Num cenário em que a já consistente investigação lexical por trás dos cinco fatores procurava integrar-se em trabalhos de pesquisadores que utilizavam questionários como ferramenta de investigação, dois pesquisadores foram responsáveis pelo desenvolvimento do Inventário de Personalidade NEO Revisto (NEO-PI-R) (Costa e McCrae, 1992), um questionário capaz de medir os cinco grandes fatores a partir de facetas específicas de cada fator (John, Naumann e Soto, 2008).

Durante a década de 1990, o modelo dos cinco grandes fatores já contava com um grande suporte empírico para provar a sua validade e um corpo de pesquisa crescente que apontava a relação entre os fatores e os resultados de interações das pessoas com o ambiente (John, Naumann e Soto, 2008). Ainda assim, alegando que o desenvolvimento científico poderia ficar subjugado ao interesse comercial devido aos direitos cobrados pelo uso dos testes, Goldberg (1999) propôs elaborar escalas com base em testes com itens teoricamente e

empiricamente válidos, além de teoricamente relevantes para a análise dos comportamentos sociais, incluindo o NEO-PI-R, porém que fossem de domínio público. Argumentando que “it is my prediction that over the next few years the phrase ‘public domain’ will come to mean accessible via the WorldWideWeb” (Goldberg, 1999, p. 11), Goldberg desenvolveu uma página na internet para disseminação do *International Personality Item Pool* (IPIP), projeto que continua em andamento e que pode ser acessado em <https://ipip.ori.org/>.

Nos anos seguintes, o conjunto inicial de 1252 itens relacionados com a personalidade foi expandido para um valor superior a 2000, além de terem sido traduzidos para 28 idiomas (Goldberg et al., 2006). Os 1252 itens iniciais do IPIP foram desenvolvidos por Lewis Goldberg e colaboradores a partir de duas fontes principais: (a) listas de traços extraídos de estudos lexicais clássicos (como Allport e Odbert, 1936) e (b) itens de instrumentos psicométricos estabelecidos (por ex., NEOPI, CPI), adaptados para domínio público. Goldberg reviu esses descritores, criando escalas que refletiam o modelo dos Cinco Grandes, visando uma ampla cobertura e uma coerência interna antes de expandir o conjunto para mais de 2000 itens em versões posteriores (Goldberg et al., 2006). Com atributos tais como gratuidade, velocidade, precisão e ausência de direitos autorais, as escalas IPIP foram amplamente utilizadas por profissionais de diversas áreas (Goldberg et al., 2006).

O desenvolvimento de instrumentos de pesquisa de informações sobre traços de personalidade entra num novo paradigma a partir de 2007. Naquele ano, David Stillwell programou a escala IPIP correspondente ao questionário NEO-PI-R (Costa e McCrae, 1992) num aplicativo do Facebook, dando início ao projeto *myPersonality*, que é tomado como um exemplo de utilização bem-sucedida de redes sociais online para pesquisa social em larga escala (Stillwell e Kosinski, 2012). A base de dados construída por meio do projeto é vista como fundamental para o então nascente campo de estudo da personalidade no contexto de *Big Data* (Lambiotte e Kosinski, 2014).

A grande adesão de empresas, instituições e pessoas às novas tecnologias computacionais e informáticas, aliada ao desenvolvimento de novos produtos e serviços, como *smartphones* e redes sociais online, fez com que aumentasse significativamente a produção de dados acerca dos mais diversos aspetos da realidade social (Mayer-Schönberger e Cukier, 2013).

Nesse contexto, é produzida uma quantidade massiva de dados em diferentes formatos e com potencial de análise por novas ferramentas computacionais – fenómeno que se convencionou chamar *Big Data*, caracterizado tradicionalmente pelos chamados 3v: volume, velocidade e variedade (TechAmerica Foundation’s Federal Big Data Commission, 2012; Gandomi e Haider, 2015; Chen e Wojcik, 2016).

Uma das características que define o *Big Data* é que a sua abordagem exige tecnologia e técnicas mais avançadas do que os procedimentos tradicionais de análise de dados até então disponíveis (TechAmerica Foundation's Federal Big Data Commission, 2012; Song et al., 2013; Chen e Zhang, 2014). As metodologias desenvolvidas para realizar a análise destes novos conjuntos de dados recebem o nome genérico de *Big Data Analytics* (BDA); por outras palavras, técnicas e tecnologias de análise de *Big Data* (Srvanthi e Reddy, 2015), com aplicações em áreas como administração pública, ciência, comunicação, economia, educação, marketing, entre outros (Chen e Zhang, 2014; Kim, Trimi e Chung, 2014; Srvanthi e Reddy, 2015; Harlow e Oswald, 2016; Kosinski e Behrend, 2017).

Sobre a presença da psicologia enquanto ciência no âmbito de *Big Data Analytics* pode-se mencionar a publicação de guias e manuais de pesquisa em psicologia utilizando Big Data desde meados da década de 2010 (Kosinski et al., 2015; Guzzo et al., 2015; Owen e Imel, 2016; Landers et al., 2016; Lai et al., 2017; Murphy, 2017), sendo que em 2016 autores como Harlow e Oswald apontavam para a incipiência de artigos de psicologia nas principais publicações sobre o assunto.

Atualmente, no entanto, uma busca no Google Académico¹ que use as palavras-chave “psychology” + “personality” + “big data analysis” vai encontrar cerca de 7270 resultados, sendo que 6780 destes resultados foram publicados a partir do ano de 2016 em diante, o que permite localizar a ampliação do interesse no campo nos últimos 10 anos.

Neste trabalho, elegemos o caso do projeto *myPersonality* para análise, pois este foi pioneiro a associar a teoria dos traços de personalidade, hegemónica na psicologia *mainstream*, às tecnologias de BDA, mais ainda por trabalhar com dados recolhidos por intermédio das redes sociais online. O projeto *myPersonality* foi um aplicativo do Facebook lançado em 2007 que permitia que os utilizadores participassem voluntariamente em pesquisas de carácter psicológico. Utilizando um questionário de personalidade baseado na escala IPIP, os participantes recebiam *feedback* sobre os seus traços e podiam, opcionalmente, ceder os seus dados para a pesquisa académica. Criado por David Stillwell e Michal Kosinski, o projeto expandiu-se, contribuindo para dezenas de estudos publicados em periódicos, impulsionando o avanço do conhecimento sobre psicologia, comportamento humano e privacidade digital (Stillwell e Kosinski, 2015).

A base de dados *myPersonality* foi constituída a partir das informações voluntariamente cedidas por usuários do aplicativo, que completaram

1 Dados obtidos numa pesquisa realizada a 16-04-2025.

questionários de personalidade no Facebook. Cerca de 40% dos participantes optaram por partilhar os seus dados, possibilitando a recolha de informações comportamentais, linguísticas e demográficas de mais de 6 milhões de perfis. Esses dados foram utilizados para investigar correlações entre traços de personalidade e comportamentos online, impulsionando pesquisas na área de psicologia digital e contribuindo para uma compreensão de padrões comportamentais em ambientes virtuais (Stillwell e Kosinski, 2015).

De acordo com Stillwell e Kosinski (2015), ao longo dos cinco anos de existência do aplicativo *myPersonality* 40% das mais de seis milhões de pessoas que responderam aos questionários fornecidos consentiram em, de forma anónima, dar acesso a todas as informações vinculadas à sua página do Facebook, desde data de nascimento e número de amigos a conteúdo das mensagens publicadas na linha do tempo. Esses dados ofereceram uma fonte com múltiplas informações sobre os comportamentos das pessoas em sociedade, que também foram usadas para testar hipóteses sobre a manifestação dos fatores de personalidade. A base de dados construída apresentou um potencial heurístico para análise social dos fatores de personalidade dos utilizadores ao invés de uma idealização de si (Back et al., 2010) e permitiu que estes fossem acedidos tanto por meio do que se chamam “afirmações de identidade” quanto por meio do que se tem chamado “resíduos comportamentais” identificáveis nos demais dados disponíveis nos perfis dos utilizadores (Gosling et al., 2002).

Por “afirmações de identidade” compreendem-se as declarações simbólicas de um indivíduo sobre si mesmo, já por “resíduos comportamentais” compreendem-se as pistas sobre si mesmo deixadas inadvertidamente por um indivíduo (Gosling et al., 2002). Esse modelo de classificação foi construído a partir de contextos sociais nos quais as relações são presenciais, porém pode ser generalizado para ambientes virtuais (Gosling, Gaddis e Vazire, 2007). No contexto de redes sociais online, as chamadas “afirmações de identidade” e os “resíduos comportamentais” manifestam-se a partir de atributos de diferentes naturezas da rede social em questão. Afirmções de identidade podem ser encontradas em atributos linguísticos (como atualização de *status* e campos “sobre mim”) e atributos comportamentais (como grupos no qual participa e fotos do perfil) (Kaushal e Patwardhan, 2018). Por sua vez, resíduos comportamentais podem ser encontrados em atributos estruturais, isto é, comportamentos que não são diretamente autodescritivos (tais como número de amigos, densidade da rede etc.) e atributos temporais (frequência de atualização de *status*, alteração no número de amigos ao longo do tempo, etc.) (Kaushal e Patwardhan, 2018). É interessante a constatação de que frequentemente as redes sociais oferecem também informações demográficas, como a cidade natal e de residência, a data de nascimento e o nível de escolaridade.

No âmbito do *myPersonality*, atributos demográficos foram usados tanto na caracterização da amostra quanto como variáveis-preditivas, por exemplo “cidade natal” e “cidade de residência” foram exploradas para mapear variações geográficas de traços de personalidade em larga escala (Rentfrow et al., 2013), “data de nascimento”, convertida em idade, permitiu investigar como a linguagem e demais comportamentos online mudam ao longo do ciclo de vida (Schwartz et al., 2013), e “nível de escolaridade” foi incorporado a modelos de aprendizagem de máquina para aprimorar a predição de traços, atuando como variável de controlo em estudos de personalidade automatizada (Kosinski et al., 2014; Kaushal e Patwardhan, 2018).

Por meio do Facebook, o projeto *myPersonality* conseguiu a base de dados mais representativa da população geral² em termos de género, idade e região geográfica (Stillwell e Kosinski, 2012) e a mais popular para estudos sobre como os fatores de personalidade podem ser expressos nos comportamentos sociais online (Lambiotte e Kosinski, 2014). A partir dessa base de dados foram realizados estudos abrangendo temas como: variações geográficas de traços de personalidade, redes sociais, privacidade, previsão de traços individuais, felicidade (Lambiotte e Kosinski, 2014), entre outros. Além disso, trabalhos provindos da base de dados *myPersonality* contribuíram para a reformulação da política de privacidade do Facebook, bem como de leis de privacidade nos Estados Unidos e no Reino Unido (*myPersonality Project*, 2018).

Em maio de 2018, o projeto *myPersonality* foi encerrado, sendo divulgado na sua página na internet que manter a base de dados, aprovar os projetos de pesquisa, responder a perguntas e acatar regulamentos legais se tornara trabalhoso demais para os dois fundadores (*myPersonality Project*, 2018). Dois fatos ocorridos no mesmo ano contribuem para que a discussão ética acerca da existência, manutenção e fins da base de dados se aprofunde: o primeiro deles, um escândalo relacionado com o Facebook no qual os dados de mais de 80 milhões de pessoas foram obtidos ilegalmente para o desenvolvimento de tecnologias de marketing direcionado (Hern, 2018). Sendo que a ciência utilizada nessa iniciativa foi desenvolvida a partir do projeto *myPersonality* (Andrews, 2018). O segundo envolve o próprio projeto *myPersonality* e a descoberta de uma chave e senha de acesso para o banco de dados que permaneceram

2 A amostra *myPersonality* foi composta por mais de seis milhões de utilizadores do Facebook de 44 países com pelo menos 1000 respondentes cada, representando não só os Estados Unidos (≈58% da amostra), mas também nações da Europa (Reino Unido, Alemanha, Países Baixos, Espanha, França), América do Norte (Canadá), Oceania (Austrália, Nova Zelândia), Ásia (Índia, Filipinas) e América Latina (Brasil), entre outras regiões, o que garantiu uma diversidade geográfica 2,5 vezes maior do que a de amostras tradicionais (Stillwell e Kosinski, 2012).

disponíveis online nos últimos quatro anos (Waterfield e Revell, 2018). Portanto, é impossível dizer que apenas as pessoas cujos projetos de pesquisa foram aprovados tiveram acesso a dados sensíveis de milhões de pessoas.

Além disso, os próprios proponentes do projeto identificaram diversas questões éticas associadas à recolha e utilização dos dados dos participantes. Entre as principais preocupações estão os riscos relacionados com a privacidade, originados pela discrepância entre o contexto em que as informações são partilhadas e aquele em que são utilizadas para pesquisa. Adicionalmente, desafios como a manutenção do consentimento informado, a segurança dos dados e o potencial para a manipulação comportamental foram destacados, enfatizando-se a necessidade de práticas transparentes e de regulamentações específicas para proteger os direitos dos indivíduos na era digital (Stillwell e Kosinski, 2015). Os resultados do projeto *myPersonality* demonstram que a incorporação do *Big Data* no estudo da personalidade pode revolucionar a forma como são entendidos os traços de personalidade ao integrar métodos tradicionais de psicomетria com técnicas computacionais avançadas. Ao mesmo tempo em que a análise dinâmica e em larga escala de padrões comportamentais permite ampliar as análises que envolvem personalidade, o uso intensivo de dados digitais impõe desafios éticos e metodológicos que pedem uma cuidada reflexão sobre consentimento, privacidade e a responsabilidade social na condução de pesquisas nesse novo cenário digital.

Uma vez contextualizado o estudo da personalidade na psicologia e a sua relevância para a análise do comportamento social, na era do *Big Data* e no projeto *myPersonality*, pode-se agora retomar as perguntas que se visou responder na presente pesquisa: (i) como, e (ii) para que fins foram estudados os fatores sociais nos quais a personalidade se manifesta no contexto de *Big Data* em pesquisas que analisaram dados da base *myPersonality*? Para isso, realizámos uma análise de conteúdo de artigos que relatam pesquisas que analisaram dados provenientes da base do projeto *myPersonality*. Especificamente, seleccionámos artigos com relatos de pesquisas produzidas a partir de dados disponibilizados pela base *myPersonality* e procedemos à identificação dos seus objetivos, populações, alvos de investigação e suas variáveis correspondentes, as ferramentas metodológicas utilizadas, as aplicações declaradas dos resultados obtidos, e, finalmente, as considerações éticas relatadas.

Embora outros estudos tenham investigado a avaliação de personalidade nas redes sociais (Reis, 2018; Waheed et al., 2017; Kaushal e Patwardhan, 2018; Ihsan e Furnham, 2018), o seu foco tecnológico e psicométrico não substitui a necessidade de um debate ético abrangente, na psicologia e nas ciências sociais, sobre essa questão. Como foi visto, desde 2016 essas áreas têm ampliado o seu olhar para o *Big Data Analytics*, reconhecendo a migração de questionários

tradicionais para observações empíricas em larga escala (Harlow e Oswald, 2016; Lambiotte e Kosinski, 2014). Além disso, aumentaram os alertas sobre desafios éticos na recolha e no uso de dados digitais (Kaplan, Riley e Mabry, 2014; Chen e Wojcik, 2016; Harlow e Oswald, 2016).

Esta pesquisa desemboca no debate da análise social ao examinar, a partir do contexto brasileiro, implicações éticas e práticas de novas formas de avaliação dos fatores sociais nos quais a personalidade se manifesta. Ao integrar reflexões sobre segurança, privacidade e responsabilidade profissional, pretende-se orientar, frente aos riscos do uso massivo de *Big Data*, futuras políticas e práticas na psicologia social em específico e na área interdisciplinar da análise social em geral.

PROCEDIMENTOS METODOLÓGICOS

Os proponentes do projeto *myPersonality* mantêm uma página oficial na internet³ na qual apresentam o histórico do projeto e uma lista de 57 publicações,⁴ por eles selecionadas, que relatam pesquisas que foram realizadas a partir da sua base de dados. A presente pesquisa toma esse universo de 57 artigos como fonte de dados e analisa-os para compreender como e para que fins os fatores sociais nos quais a personalidade se manifesta foram estudados nessa iniciativa, que foi uma das pioneiras no contexto de *Big Data*.

Trabalhou-se com a técnica da análise de conteúdo, que, segundo Bardin (1977), visa captar conteúdos comuns e nuances de significação, cumprindo funções exploratórias e de verificação de hipóteses, bem como inferir conhecimentos sobre as condições de produção e receção dos textos.

De acordo com Bardin (1977), a análise de conteúdo tem a intenção de, a partir de indicadores quantitativos ou qualitativos presentes no texto, inferir conhecimentos sobre as suas condições de produção e de receção, para conhecer o que levou a tal enunciação e quais as suas consequências.

Seguindo o método proposto por Bardin (1977), a análise do conteúdo dos 57 artigos fundamentados na base de dados *myPersonality* foi realizada em três fases: (1) pré-análise; (2) exploração do material; (3) tratamento dos resultados, inferência e interpretação. A pesquisa foi, assim, realizada em três etapas alinhadas com o objetivo geral. Na pré-análise, foram selecionados os artigos que constavam na sessão “publicações selecionadas” do *site*, sendo o critério de seleção adotado o seguinte: tratar-se de uma pesquisa empírica que utilize o *score* da escala IPIP, correspondente ao modelo dos cinco grandes

3 Ver www.myPersonality.org.

4 Ver <https://sites.google.com/michalkosinski.com/mypersonality/publications?authuser=0>.

fatores, com pelo menos uma das variáveis digitais para o seu alvo de pesquisa, o que resultou na seleção de 23 artigos. Em seguida, na exploração do material, procedeu-se à leitura flutuante que levou à formulação de sete índices temáticos que guiaram a categorização dos conteúdos: (i) objetivo da pesquisa, (ii) população, (iii) alvos de investigação, (iv) variáveis digitais, (v) ferramentas metodológicas, (vi) aplicações declaradas e (vii) considerações éticas. Por fim, no tratamento dos resultados, as categorias foram cruzadas e organizadas em gráficos e tabelas, para então se inferir a forma como as práticas observadas respondem às perguntas “como” e “para que fins” os fatores sociais nos quais a personalidade se manifesta foram estudados no contexto do projeto.

Apresentam-se e discutem-se a seguir os resultados encontrados. No processo de inferência procurou-se conectar propriedades formais dos artigos (os sete índices temáticos) com as suas implicações acadêmicas e sociais, de forma a descrever os padrões encontrados na pesquisa concomitantemente à discussão das suas consequências sociais e impactos éticos.

RESULTADOS E DISCUSSÃO

Dos 57 artigos disponíveis na base de dados original, foram selecionados 23 de acordo com os critérios especificados anteriormente. Sobre os artigos analisados foi possível identificar: que foram publicados no período de 2010 a 2018, tiveram 13 alvos de investigação diferentes, usaram 73 variáveis digitais correlacionadas. Essas investigações usaram como metodologia de análise 52 ferramentas metodológicas diferentes, variando desde procedimentos comuns, como coeficiente de correlação de Pearson, a algoritmos mais avançados e passando por softwares de análise de linguagem. A presente secção está estruturada da seguinte maneira: inicia-se a apresentação dos resultados listando as referências analisadas, os índices temáticos (i) “objetivos” e (ii) “número de participantes” de cada estudo, para que se tenha um panorama dos artigos. Em seguida, em subsecção própria, é apresentado o índice temático (iii) “alvos de investigação” de cada artigo, para que estes possam ser discutidos sob uma perspectiva mais ampla. Na sequência, é feita a apresentação dos índices temáticos (iv) “variáveis digitais” e (v) “ferramentas metodológicas”, numa mesma subsecção, para que se possa responder de que forma os fatores de personalidade têm sido estudados. Por fim, a última subsecção é dedicada a apresentar e discutir os índices temáticos (vi) “aplicações” e (vii) “considerações éticas”.

QUADRO 1

Apresentação dos artigos analisados

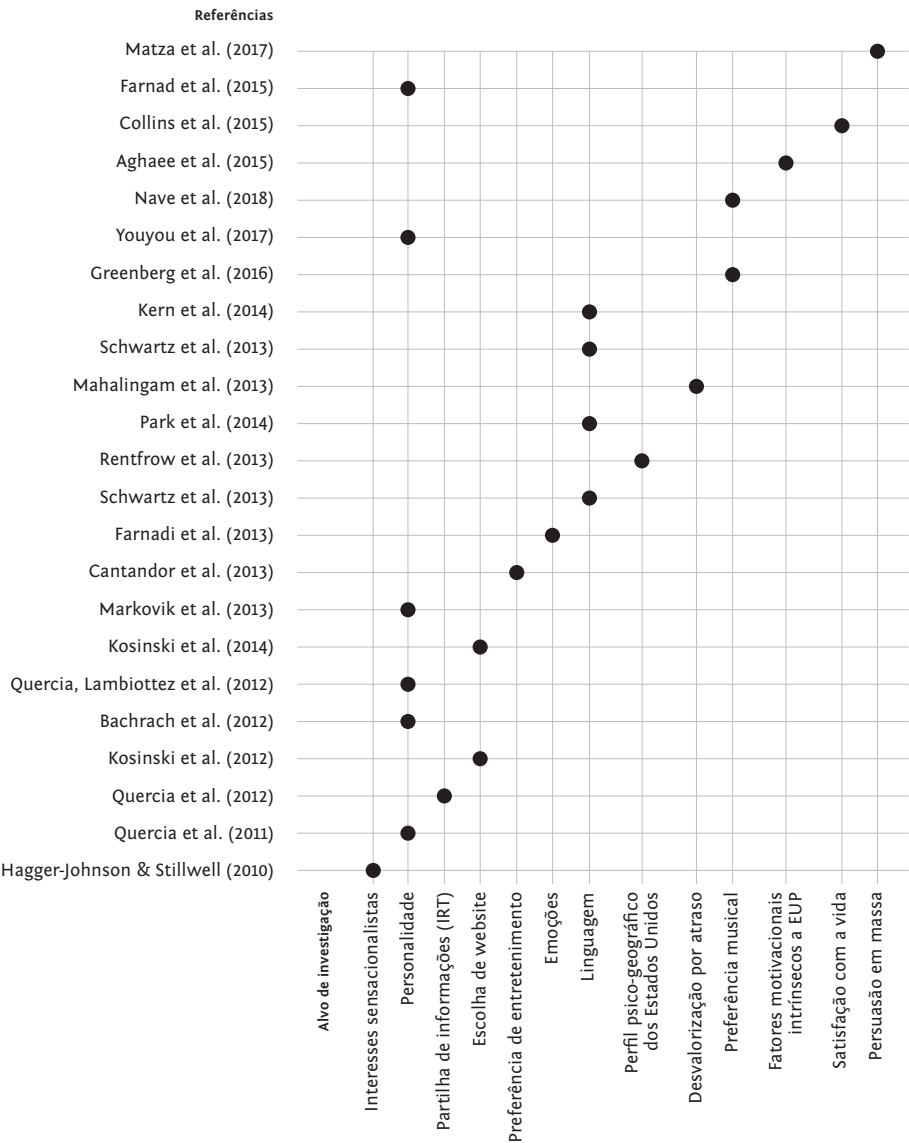
Referência	Título	Objetivo	#Participantes
Hagger-Johnson, Egan e Stillwell (2010)	Are Social Networking Profiles Reliable Indicators of Sensational Interests?	Investigar se informações postadas online são indicadores confiáveis de interesses sensacionalistas	694
Quercia, Kosinski e Stillwell, Crowcroft (2011)	Our Twitter Profiles, Our Selves: Predicting Personality with Twitter.	Analisar a relação entre personalidade e diferentes tipos de utilizadores do Twitter	335
Quercia <i>et al.</i> (2012)	Facebook and Privacy: The Balancing Act of Personality, Gender, and Relationship Currency.h	Analisar a relação entre personalidade e revelação de informações	1313
Bachrach, Kosinski, Graepel, Kohli e Stillwell (2012)	Personality and Patterns of Facebook Usage.	Demonstrar como a personalidade se manifesta em diferentes atributos do Facebook	180 000
Kosinski, Stillwell, Kohli, Bachrach e Graepel (2012)	Personality and Website Choice.	Analisar a relação entre personalidade e preferência de websites	10 897
Quercia, Lambiotte, Stillwell, Kosinski e Crowcroft (2012b)	The Personality of Popular Facebook Users.	Analisar a relação entre popularidade e traços de personalidade	172 952
Cantador, Fernández-Tobías, Bellogín, Kosinski, Stillwell (2013)	Relating Personality Types with User Preferences in Multiple Entertainment Domains.	Entender a relação entre preferência por tipos de entretenimento e personalidade	53 226
Markovikj, Gievska, Kosinski e Stillwell (2013)	Mining Facebook Data for Predictive Personality Modeling.	a) Identificar indicadores relevantes relacionados com a personalidade contidos explícita ou implicitamente nos dados de utilizadores do Facebook; b) Investigar a falseabilidade de modelos preditivos da personalidade para apoiar sistemas inteligentes	250
Schwartz <i>et al.</i> (2013)	Toward Personality Insights from Language Exploration in Social Media.	Identificar palavras correlacionadas com a personalidade	74,941
Rentfrow, Gosling, Jokela, Stillwell e Kosinski (2013)	Divided We Stand: Three Psychological Regions of the United States and their Political, Economic, Social, and Health Correlates.	Construir um mapa dos Estados Unidos baseado na personalidade	1 596 704
Mahalingam, Stillwell, Kosinski, Rust e Kogan (2013)	Who Can Wait for the Future? A Personality Perspective.	a) Existem diferenças na personalidade relacionadas com o engajamento e a desvalorização pelo atraso? b) Como é que as diferenças de personalidade afetam a magnitude do efeito?	a) 5909 b) 5888
Schwartz, Eichstaedt <i>et al.</i> (2013)	Personality, Gender, and Age in the Language of Social Media: The Open-vocabulary Approach.	Explorar atributos linguísticos nas redes sociais em função da idade, género e personalidade	74 941

Referência	Título	Objetivo	#Participantes
Farnad et al. (2014)	How Are You Doing? Emotions and Personality in Facebook.	Especificar a relação entre as emoções dos utilizadores e outras características do Facebook	5865
Kern et al. (2014)	The Online Social Self: An Open Vocabulary Approach to Personality.	Apresentar um novo método de vocabulário aberto para analisar as diferenças individuais	69 792
Park et al. (2014)	Automatic Personality Assessment Through Social Media Language.	Acessar a personalidade automaticamente através da linguagem utilizada na rede social	66 000
Kosinski, Bachrach, Kohli, Stillwell e Graepel (2014)	Manifestations of User Personality in Website Choice and Behaviour on Online Social Networks.	Examinar como a personalidade se manifesta em comportamentos online e é refletida pela escolha de sites	350 000
Aghaee, Blackwell, Kosinski e Stillwell (2015)	Personality and Intrinsic Motivational Factors in End-User Programming.	Identificar fatores intrínsecos à motivação para programação pelo utilizador final	100
Collins, Sun, Kosinski, Stillwell e Markuzon (2015)	Are You Satisfied with Life?: Predicting Satisfaction with Life from Facebook	Identificar satisfação com a vida; eficiência e eficácia	101 069
Farnad et al. (2016)	Computational Personality Recognition in Social Media.	Comparar os métodos mais avançados de análise computacional da personalidade	4237
Greenberg et al.(2016)	The Song Is You: Preferences for Musical Attribute Dimensions Reflect Personality	a) identificar a estrutura por trás de atributos musicais percebidos b) identificar a relação entre preferência musical e personalidade	a) 76 b) 9454
Youyou, Stillwell, Schwartz, Kosinski (2017)	Birds of a Feather Do Flock Together: Behavior-Based Personality-assessment Method Reveals Personality Similarity among Couples and Friends.	Examinar o grau da semelhança de personalidade entre casais românticos e entre amigos	138 553
Matz, Kosinski, Navec e Stillwell (2017)	Psychological Targeting as an Effective Approach to Digital Mass Persuasion	Testar os efeitos da persuasão psicológica num ambiente ecologicamente válido	58 466
Nave et al. (2018)	Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes	Investigar se os traços de personalidade se manifestam no comportamento de ouvir música	22 252

ALVOS DE INVESTIGAÇÃO

Procurou-se analisar o foco de investigação dos estudos, fazendo referência à variável ou constructo psicológico que foi focado por cada um dos artigos a partir da descrição do objetivo do artigo, conforme disposto no Quadro 1. Essas informações são apresentadas de forma sistematizada na Figura 1.

FIGURA 1
O alvo de investigação para cada referência



Foram identificados 13 alvos de investigação diferentes, sendo “personalidade” e “linguagem” os mais comuns, aparecendo respectivamente em 6 e 4 artigos cada. Outros alvos de investigação que se repetiram foram “preferência por *website*” e “preferência musical”, ambas com duas aparições. Assim, a mesma variável ou constructo psicológico pode ter sido estudada a partir de diferentes facetas ou de diferentes variáveis digitais.

Para se discutir as variáveis identificadas é importante retomar o contexto geral em que esses dados são apresentados. Ou seja, ir além das justificações presentes em cada artigo e pensar essa escolha como parte integrante de um corpo de pesquisa amplo, o que acontece num contexto social e económico específico e considerar que, em 2017, 70% dos norte-americanos, que são a maioria da população alvo dos artigos analisados, navegavam na internet; 59% ouviam/descarregavam músicas online; e 36% viam/descarregavam vídeos online diariamente (Center for the Digital Future, 2017).

Considere-se também que recolher informações sensíveis dos hábitos online é possível tecnicamente, por meio, por exemplo, do uso de *cookies* (Mayer e Mitchell, 2012). Assim, escolher estudar a relação entre personalidade e preferência por websites (como fazem Kosinski et al., 2012; Kosinski et al., 2014), preferência musical (como fazem Greenberg et al., 2016; Nave et al., 2018) ou preferência por entretenimento (como fazem Cantador et al., 2013) permite inferir um interesse em identificar o modo como a personalidade dos internautas se manifesta a partir das suas atividades mais frequentes e sem a necessidade de recorrer ao Facebook, como no caso dos artigos analisados. A questão contextual fica ainda mais explícita quando se considera que, por exemplo, Bachrach e colaboradores (2012), cujo trabalho investigou a preferência por website, são pesquisadores da Microsoft, o que pode ser indicativo de um interesse daquela instituição pela temática.

É também importante considerar que até mesmo a ausência de informações pode ser tomada como indicador de traços de personalidade. Tal como a análise da relação entre a personalidade e a partilha de informações realizada por Quercia e colaboradores (2012a), que encontraram uma relação entre traços de personalidade, informações disponibilizadas (e não disponibilizadas) e género.

Na hipótese de não haver dados disponíveis sobre um utilizador, infere-se que há relação de similaridade entre a sua personalidade e a personalidade dos seus amigos e par romântico (Youyou et al., 2017). Portanto, o conjunto das variáveis-alvo escolhidas revelou uma procura por formas de aprimorar a identificação de manifestação/expressão dos fatores/dimensões da personalidade em conjunto com informações demográficas, bem como relacioná-las com hábitos frequentes, forma de se expressar (linguagem) e outros

constructos psicológicos, como satisfação com a vida (Collins et al., 2015) e interesses sensacionalistas (Hagger-Johnson, Egan e Stillwell, 2011).

Compreende-se, assim, depois de se analisar o foco de investigação dos estudos, fazendo referência às variáveis psicológicas que foram focadas, que, além de revelar que dimensões do constructo “personalidade” foram alvo de modelagem preditiva, esses estudos apontam o modo como tais traços se expressam em interações sociais mediadas por plataformas digitais. Ao correlacionar atividades online, tais como *likes*, partilhas, padrões de linguagem e ausência de postagens, com as cinco dimensões da personalidade conforme o modelo *Big Five* demonstra-se que índices de “extroversão”, “amabilidade”, “conscienciosidade”, “neuroticismo” e “abertura” não são analisados a partir de atributos isolados, mas a partir de manifestações em práticas sociais concretas: na formação de redes de amizade, no engajamento em comunidades, na recomendação de conteúdos e na gestão da privacidade, inserindo a análise da personalidade no âmbito da sociabilidade digital, mostrando que “assinaturas” psicométricas capturadas em *Big Data* representam padrões de interação social que refletem e reforçam posicionamentos, normas e vínculos coletivos em ambientes virtuais.

VARIÁVEIS DIGITAIS E FERRAMENTAS METODOLÓGICAS

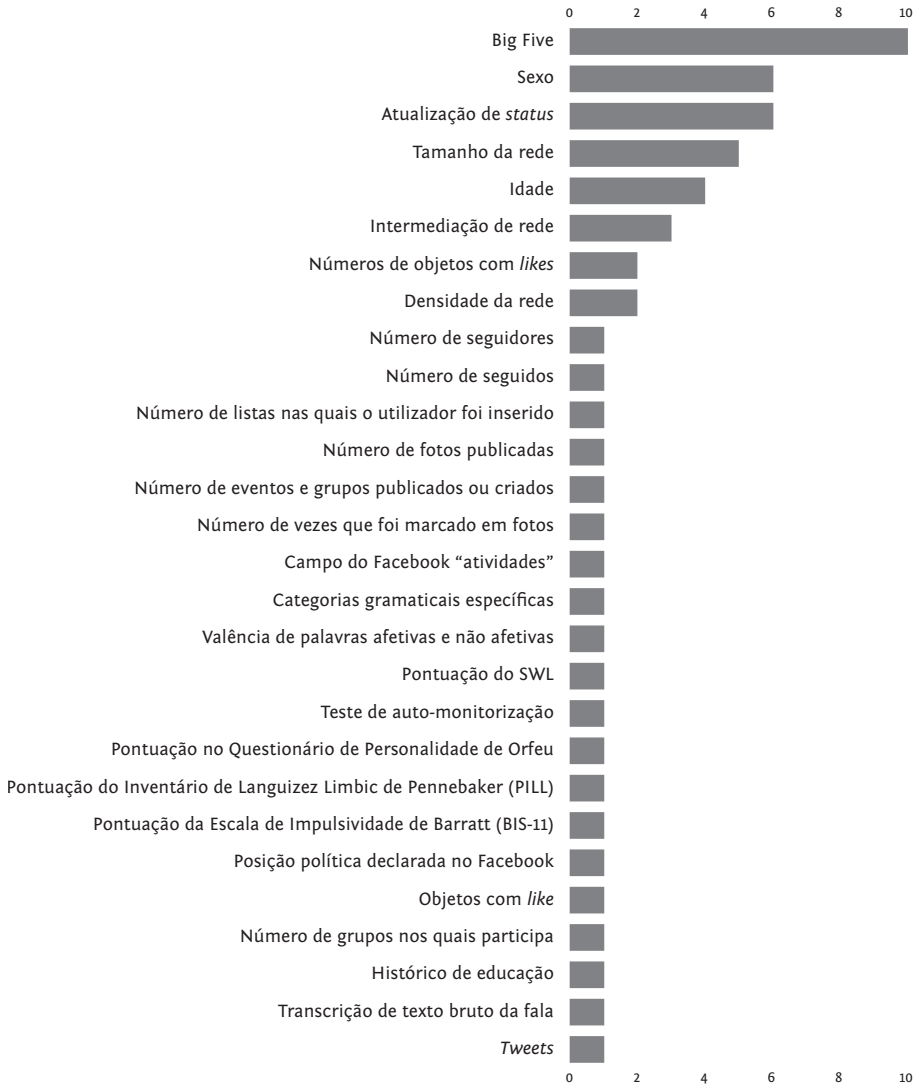
Procedeu-se, em seguida, à análise de que variáveis digitais foram consideradas nos estudos bem como de que ferramentas metodológicas foram utilizadas. Os resultados permitem destacar o impacto da introdução dessas novas tecnologias na psicometria.

Procurou-se identificar e sistematizar as variáveis digitais correlacionadas com os alvos de investigação dos artigos analisados, ou seja, compreender a partir de que dados disponíveis na base *myPersonality* foram conduzidos os estudos. As variáveis que se repetiram mais vezes foram: sexo (12), idade (11), tamanho da rede (9), atualização de *status* (6), número de objetos que receberam *likes* (5). Em seguida, número de marcação de fotos, densidade da rede e número de grupos no qual participa contam com 3 aparições cada. Além destas, outras 8 variáveis surgiram duas vezes, deixando assim um total de 57 variáveis utilizadas apenas uma vez. Conforme se pode visualizar na Figura 2, os estudos que têm como alvo de investigação a personalidade e a linguagem abrangem um total de 28 variáveis digitais.

Procurou-se identificar e sistematizar que ferramentas ou técnicas diferentes foram empregadas na análise dos dados de cada artigo. O objetivo desta análise foi tornar conhecidos os nomes de softwares, algoritmos e procedimentos estatísticos usados para pesquisa no contexto de *Big Data*. Foram identificadas

FIGURA 2

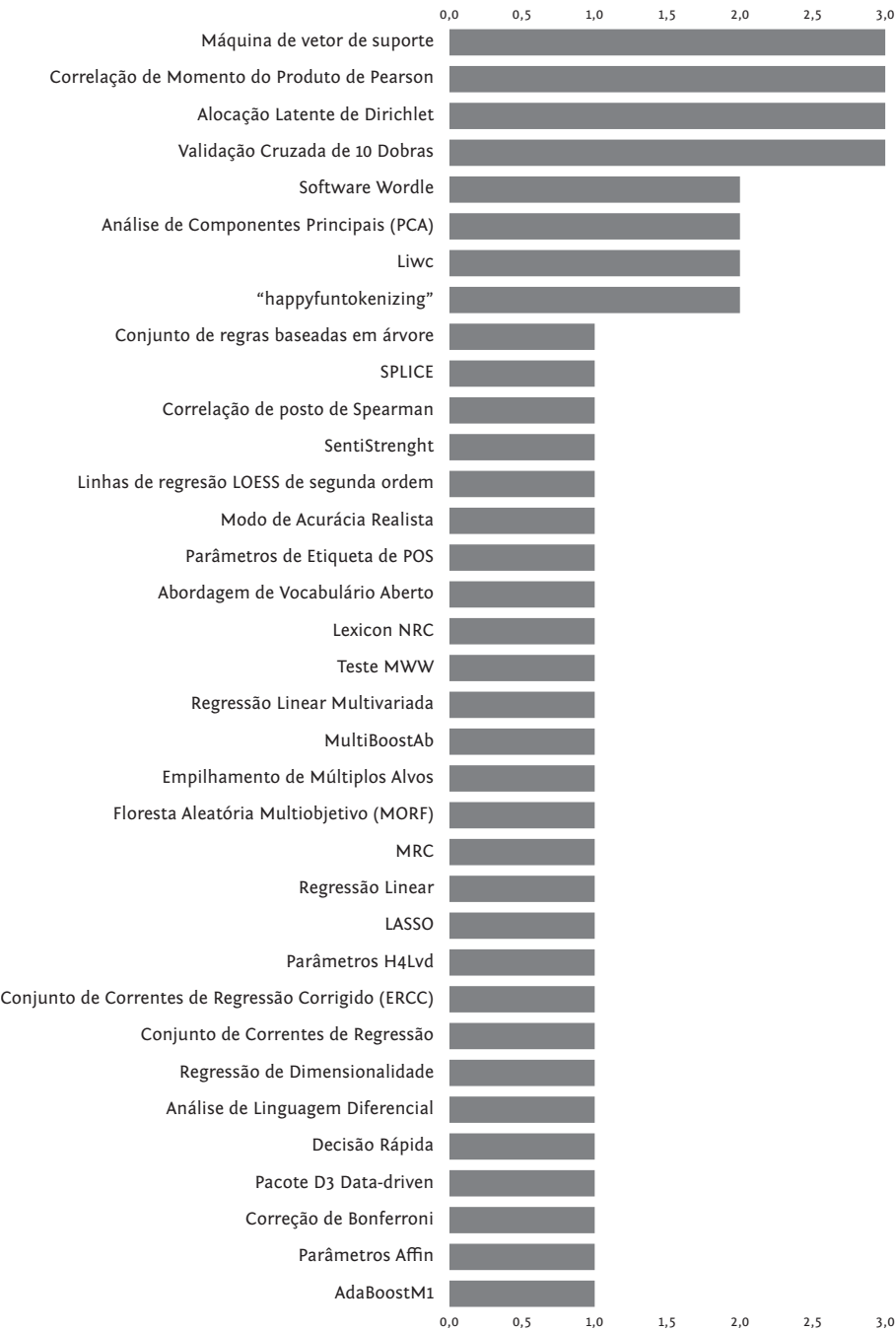
Variáveis digitais utilizadas para investigação da personalidade e da linguagem



52 ferramentas metodológicas diferentes, das quais 30 foram utilizadas para os estudos que investigam a personalidade e a linguagem, conforme se pode verificar na Figura 3.

A compreensão de que variáveis digitais e de que ferramentas metodológicas foram mais utilizadas nos trabalhos aqui analisados evidencia importantes elementos indicativos que podem ajudar à construção da resposta para a pergunta “como é que a personalidade é estudada no contexto de *Big Data*?”.

FIGURA 3
Ferramentas metodológicas utilizadas para investigação da personalidade e da linguagem



Para esse fim, convém entendê-los como uma interface entre a psicometria tradicional e a psicologia computacional.

A psicologia computacional pode ser definida, de forma ampla, como o campo de pesquisa que procura adaptar modelos cognitivos para algoritmos com o fim de tornar tais modelos executáveis em computadores (Sun, 2008). No caso do estudo da personalidade com base no modelo *Big Five*, o que é “algoritmizado” deixa de ser um constructo psicológico em si para se tornar na maneira de lhe aceder.

Por outras palavras: da análise dos trabalhos aqui selecionados, realizados a partir de dados disponibilizados na base *myPersonality*, foi possível compreender que testes psicométricos foram adaptados para serem realizados via algoritmos; esses algoritmos processam dados disponíveis online; os dados que estão disponíveis online podem ser produzidos pelos indivíduos, utilizadores das redes sociais, de forma direta (via aquilo a que se chama atributos comportamentais e linguísticos) e/ou residual (via aquilo a que se chama atributos estruturais e temporais).

Essa adaptação rompe com a tradição da prática psicométrica em termos de aplicabilidade, abrangência potencialmente massiva, *expertise* exigida e contexto de aplicação. Os algoritmos conseguem processar, num curto período de tempo, dados de centenas de milhares de pessoas de acordo com métricas e modelos específicos. Isso traz diversas implicações que merecem reflexão.

Boa parte das ferramentas metodológicas identificadas são procedimentos estatísticos implementados em algoritmos de uso comum na aprendizagem da máquina, área cujo objetivo principal é construir modelos preditivos a partir dos dados analisados (Yarkoni e Westfall, 2017).

Argumenta-se que o propósito de *Big Data* estará mais voltado para a previsão do que para a explicação privilegiando a construção de modelos preditivos capazes de processar automaticamente grandes volumes de informações para antecipar comportamentos futuros (Yarkoni e Westfall, 2017; Lin, 2015).

Já a psicologia, em termos gerais, concentrou-se em formular explicações causais, isto é, perceber porque é que certos fenómenos ocorrem, a partir de teorias testadas empiricamente. No entanto, essa ênfase explicativa, embora rica em *insights* sobre processos mentais e sociais, nem sempre gera um poder preditivo robusto (Yarkoni e Westfall, 2017).

Como observa Shmueli (2010), presume-se muitas vezes que modelos com elevado poder explicativo possuem automaticamente alto poder preditivo, mas isso raramente se verifica na prática. Um exemplo de desconexão fica explícito no *Reproducibility Project*, que tentou replicar 100 estudos clássicos de psicologia e obteve sucesso em apenas 36% dos casos, mostrando que *insights* causais não garantem previsões robustas (Open Science Collaboration,

2015). Yarkoni e Westfall (2017) reforçam que, apesar das teorias detalhadas sobre mecanismos psicológicos, o seu poder de prever comportamentos futuros costuma ser limitado, defendendo o uso de técnicas de *machine learning* para aumentar a capacidade preditiva da disciplina.

Desta forma, abre-se então espaço para o debate acerca da efetividade prática desses modelos. A transição proposta implica que a psicologia incorpore métodos de *Big Data* para aumentar a sua capacidade de previsão, sem abandonar o seu compromisso com a compreensão dos mecanismos subjacentes. Compreende-se, assim, a importância de as pesquisas em psicologia e em análise social como um todo tentarem articular explicação e predição, podendo, por exemplo, usar algoritmos de aprendizagem da máquina para gerar hipóteses e, simultaneamente, explorar essas hipóteses por meio de desenhos experimentais e teóricos que oferecem interpretação causal e validação dos modelos.

A formulação do modelo dos cinco grandes fatores de personalidade surgiu a partir da teoria dos traços, que, partindo de uma base empírica, não tinha a explicação como um objetivo de pesquisa (John, Naumann e Soto, 2008) – o modelo *Big Five* emergiu de uma investigação léxica empírica sem fundamentação teórica inicial, e teve sua consistência ampliada pelos trabalhos empíricos subsequentes (Hutz et al., 1998; Silva e Nakano, 2001). No entanto, pesquisas futuras em psicologia e análise social através de *Big Data* terão que aprofundar o debate sobre explicação *versus* previsão. Enquanto as validações de modelos teóricos sem o enfoque preditivo produziram resultados cientificamente questionáveis (Open Science Collaboration, 2015), formular perguntas de pesquisa que podem ser respondidas de forma preditiva possibilita a formulação de teorias explicativas com uma maior base científica (Yarkoni e Westfall, 2017). Num cenário em que algoritmos de *Big Data* revelam padrões de comportamento individual e coletivo com notável precisão (Yarkoni e Westfall, 2017), cabe à psicologia social, e em particular ao campo da análise social, não se contentar apenas com a predição (Harlow e Oswald, 2016).

A capacidade de prever comportamentos online ou de identificar traços de personalidade pode contribuir, mas sendo ainda insuficiente, para orientar políticas públicas, intervenções sociais ou práticas responsáveis de marketing e saúde mental (Lambiotte e Kosinski, 2014). Há uma necessidade de explicação dos mecanismos que geram esses padrões, sem a qual se corre o risco de naturalizar correlações e perpetuar vieses, como ficou demonstrado pelas dificuldades de replicação em psicologia quando o enfoque explicativo foi negligenciado (Open Science Collaboration, 2015).

Para que futuras pesquisas articulem predição e explicação é importante promover a transparência metodológica de forma a que os atores sociais, sejam

indivíduos, comunidades ou instituições, tenham acesso tanto aos resultados algorítmicos como às razões sociais, culturais e institucionais que fundamentam tais inferências (Chen e Wojcik, 2016). A falta de responsabilização ética na realização de pesquisas no contexto do *Big Data* pode estar a deixar a descoberto caminhos que a pesquisa em psicometria historicamente constituiu, permitindo que determinados modelos funcionem como “caixas-pretas” incontestáveis e como instrumentos pouco passíveis de questionamento e validação (Kaplan, Riley e Mabry, 2014).

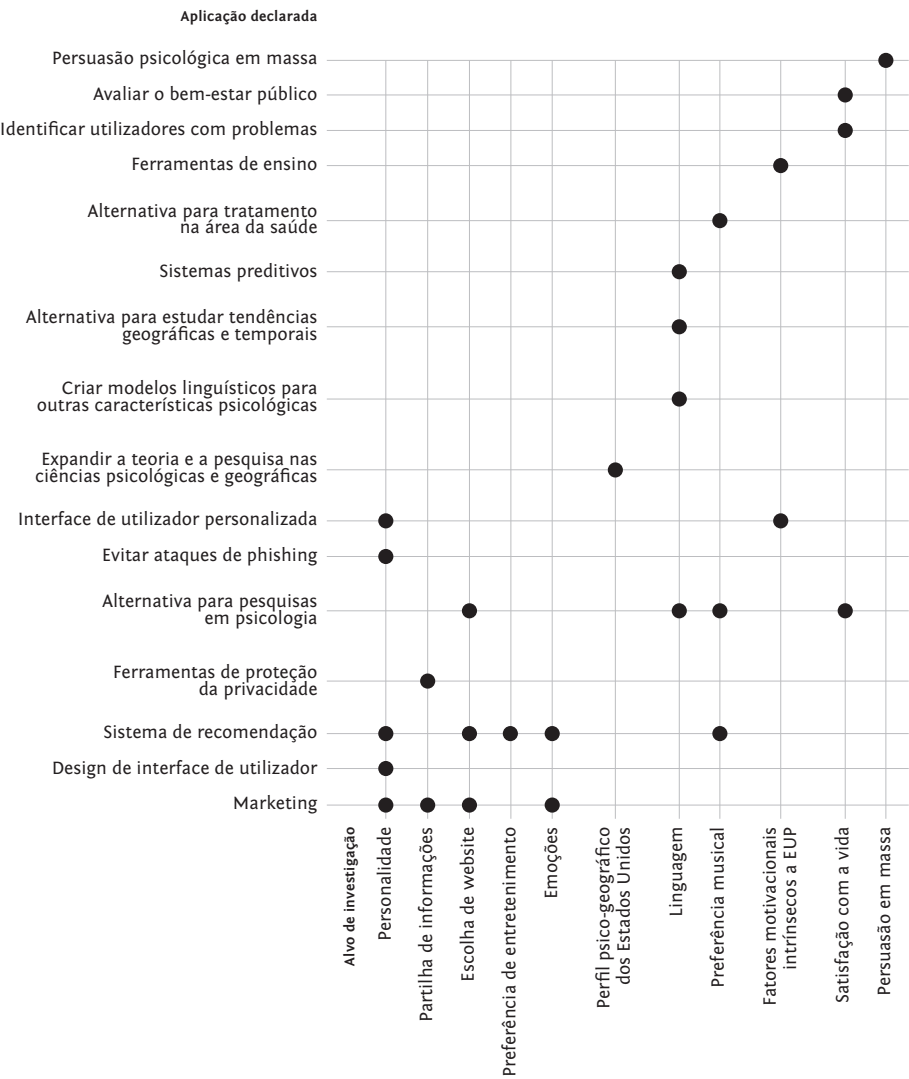
Um dos fatores aos quais a qualidade desse debate está sujeita tem que ver com quanto o ambiente acadêmico da psicologia e da análise social está apetrechado de ferramentas modernas de análise de dados. A extração de informação a partir de *Big Data* ou até mesmo a formulação de métricas que podem ser processadas algorítmicamente exigem um trabalho multidisciplinar (Chen e Zhang, 2014; Lambiotte e Kosinski, 2014).

Há, por um lado, a necessidade de os pesquisadores intensificarem a pesquisa conjunta com outras áreas do conhecimento para um melhor aproveitamento do uso de *Big Data* e um refinamento das práticas de pesquisa da comunidade acadêmica. Enquanto a pesquisa em psicologia no Brasil ainda não se mostra familiarizada com as novas tecnologias, acadêmicos de outras áreas já usam este e outros softwares de análise linguística – inclusive para o estudo da personalidade (Machado, 2016; Silva, 2018; Carvalho, Silva e Natividade, 2023). Entende-se, portanto, que está posto à psicologia e à análise social o desafio de debater modelos computacionais preditivos da personalidade sob um ponto de vista tanto técnico quanto ético.

APLICAÇÕES E CONSIDERAÇÕES ÉTICAS DECLARADAS NOS ARTIGOS

Como já foi apresentado, na análise dos 23 artigos selecionados para o presente estudo, que relataram pesquisas realizadas a partir dos dados disponíveis na base *myPersonality*, procurou-se compreender quais eram os seus alvos de investigação, as variáveis digitais e as ferramentas metodológicas utilizadas. Nesta secção, apresenta-se o que se identificou como declarado nesses trabalhos sobre que aplicações vislumbram os autores para esses estudos (sistemizado na figura 4), bem como as considerações éticas que apresentaram. Os resultados permitem destacar conflitos de interesse entre atores sociais envolvidos no processo de pesquisa (por exemplo, a quem se destina a pesquisa, quem executa a pesquisa, quem participa na definição dos objetivos e aplicações da pesquisa, e quem é tomado como objeto de pesquisa) e problemas quanto ao respeito de direitos individuais.

FIGURA 4
Relação entre alvo de investigação e aplicação declarada dos resultados



As aplicações declaradas pelos autores mais presentes nos estudos foram as seguintes: aplicação em sistemas de recomendação, marketing e alternativas de pesquisa em psicologia. Identifica-se, portanto, uma tendência para que a identificação/classificação de traços de personalidade por meio de modelos computacionais não vise benefícios diretos para os indivíduos envolvidos, mas sim motivos de governabilidade, isto é, aplicações que influenciem o comportamento das pessoas. Como exemplo, destacam-se as considerações de Matz

e colaboradores (2017) sobre a forma como a modificação de anúncios com base na análise de um único objeto que recebeu *likes* pode aumentar até 50% o número de compras de um produto. Nas palavras dos autores:

Os nossos resultados sugerem que a aplicação de direcionamento psicológico torna possível influenciar o comportamento de grandes grupos de pessoas personalizando apelos persuasivos para as necessidades psicológicas do público-alvo. (Matza et al., 2017, p. 1)

Realce-se que foram identificadas considerações éticas explicitadas pelos autores em apenas 5 dos 23 artigos analisados. Elas podem ser resumidas da seguinte maneira: o contexto no qual as pessoas optam por partilhar informações é diferente do contexto no qual essas informações serão utilizadas (Quercia et al., 2012a). Assim, quando inferências sobre informações pessoais sensíveis são retiradas a partir de dados percebidos como irrelevantes, as expectativas de privacidade podem ser frustradas devido à exploração de pontos fracos, como o vício em apostas (Kosinski et al., 2012; Quercia et al., 2012a; Quercia, et al., 2012b; Kosinski et al., 2014; Matz et al., 2017).

Portanto, respondendo à pergunta “para que fins tem sido a personalidade estudada no contexto de *Big Data*?”, a partir da análise dos estudos pioneiros que foram realizados com dados disponíveis na base *myPersonality* identifica-se que tem sido maioritariamente usada para aumentar o poder de adaptação de iniciativas de marketing e serviços face às características individuais do seu público-alvo.

Como consideração adicional sobre ética e também relacionada com conflitos de interesse entre atores sociais envolvidos no processo de pesquisa, cabe destacar ainda que, no caso específico da psicometria no contexto brasileiro, esta nova forma de analisar dados sobre constructos psicológicos no contexto do *Big Data*, especialmente, como foi visto, quando relacionado com o estudo da personalidade, a sua relação com o comportamento social e o funcionamento da sociedade entra em contradição com o entendimento dos psicólogos organizados no Conselho Federal de Psicologia. No entendimento destes atores sociais quanto ao compromisso social da psicologia (Bock et al., 2022), compreende-se que os testes psicológicos devem ser aplicados considerando os interesses e direitos das pessoas e da sociedade como um todo (Tavares, 2010).

Vale a pena recapitular que a construção de modelos computacionais preditivos da personalidade surgiu com a criação de um modelo do International Personality Item Pool que visava reproduzir estatisticamente o NEO-PI-R (Costa e McCrae, 1992). A versão brasileira do NEO-PI-R é um instrumento regulamentado e considerado de uso exclusivo do psicólogo no Brasil (Tavares, 2010).

Uma indagação colocada por Tavares (2010) sintetiza dilemas éticos e políticos emergentes quando as pesquisas sobre constructos psicológicos começam a ser realizados no contexto de *Big Data*: até que ponto a responsabilidade de proteger os direitos e os interesses dos indivíduos se aplica não apenas a profissões como a psicologia, e a instrumentos formais e regulados como os testes psicológicos, mas também a algoritmos e a modelos preditivos que inferem traços psicológicos sem intervenção humana direta?

Ao automatizar avaliações de personalidade, criam-se fronteiras de poder e de vulnerabilidade: empresas, governos e pesquisadores podem usar essas inferências para fins comerciais, de vigilância (Simões e Augusto, 2019) ou de influência política (Messenberg e Camargos, 2024; Rutledge, 2018), sem que os sujeitos tenham conhecimento ou controlo sobre o modo como os seus dados são processados.

A partir do referencial do compromisso social do psicólogo no Brasil (Bock et al, 2022) compreende-se que esse cenário exige, portanto, não apenas avanços metodológicos, mas também documentos de regulação e debates públicos que garantam transparência, consentimento efetivo e mecanismos de responsabilização, de modo a equilibrar o potencial inovador da pesquisa psicométrica com *Big Data* com a salvaguarda dos direitos individuais e da justiça social.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Nesta pesquisa foi realizada a análise de conteúdo de publicações selecionadas do projeto *myPersonality* relativas ao tema personalidade com o objetivo de compreender como e para que fins são estudados os fatores sociais nos quais a personalidade se manifesta no contexto de *Big Data*. Quanto ao “como”, foi possível concluir que a manifestação dos fatores de personalidade tem sido estudada através da adaptação de testes psicométricos de traços de personalidade já clássicos na psicologia hegemónica a algoritmos que processam dados disponíveis online, que são produzidos pelos indivíduos de forma direta (atributos comportamentais e linguísticos) e/ou residual (atributos estruturais e temporais). Testes clássicos (por ex., IPIP/NEO-PI-R) foram convertidos em rotinas de *machine learning*, como regressão, SVM e redes neurais, que consideram desde “likes”, textos publicados, horários e padrões de conexão para gerar *scores* de traços em escala contínua, validando-os por correlação com medidas tradicionais e otimizando a predição via *cross-validation*. Quanto a “para que fins”, a principal finalidade expressada nos artigos analisados é fornecer tecnologia para aumentar o poder de adaptação de empresas de marketing e serviços face às características individuais do seu público-alvo.

Os estudos revelaram que cada um dos Cinco Grandes Fatores de Personalidade foram traduzidos em padrões sociais digitais mensuráveis: “extroversão” aparece na amplitude e frequência das interações e tamanho das redes de amizade (Bachrach et al., 2012); “abertura” expressa-se pela riqueza vocabular e pelo engajamento em conteúdos culturais (Schwartz et al., 2013); “conscienciosidade” associa-se à regularidade temporal das postagens e ao uso de termos de planejamento (Park et al., 2015); “amabilidade” manifesta-se em comportamentos colaborativos, como marcações em grupo (Youyou et al., 2017); e “neuroticismo” em indicadores linguísticos de aflição e autorreferência sob stress (Farnadi et al., 2014). Esses resultados confirmam que os traços de personalidade são incorporados nas práticas sociais online, sendo posteriormente tomados como instrumento para mapear dinâmicas de sociabilidade, coesão e conflito em ambientes digitais, com implicações para intervenções sociais, marketing personalizado e formulação de políticas públicas.

O trabalho evidenciou que a avaliação de personalidade no contexto do *Big Data* traz dilemas éticos, especialmente no que tange ao consentimento informado dos sujeitos estudados, à transparência dos algoritmos e à proteção da privacidade dos indivíduos. Ao demonstrar como dados secundários, sejam linguísticos, comportamentais e/ou demográficos, são reutilizados para inferir traços psicológicos sem o pleno conhecimento dos sujeitos realça-se também o risco de manipulação comportamental decorrente do uso desses dados por parte de empresas e de instituições. Argumenta-se, portanto, que, para além dos avanços metodológicos, é imprescindível estabelecer marcos regulatórios, práticas de governança de dados e mecanismos de responsabilização que garantam o uso justo e responsável dessas inferências, alinhando a inovação em *Big Data* com os princípios éticos da pesquisa com seres humanos. Além de respeitar a Lei Geral de Proteção de Dados Pessoais (LGPD) (Brasil, Lei n.º 13.709/2018) e o *General Data Protection Regulation (GDPR)* (EU, Regulamento n.º 2016/679), é urgente exigir avaliações de impacto de privacidade específicas para a inferência de traços psicológicos, com consentimento explícito e direito de explicação sobre algoritmos, tais como normas que tornem obrigatórias as auditorias independentes de “caixas-pretas” algorítmicas e estabeleçam sanções para o uso discriminatório ou manipulativo desses perfis.

Este estudo visa contribuir para demonstrar empiricamente como os Cinco Grandes Fatores de Personalidade têm sido identificados por meio de atributos digitais (linguagem, rede, tempo), ampliando a compreensão da psicometria no contexto de *Big Data*, apontando reflexões metodológicas, éticas e políticas sobre essas práticas. Ao converter testes psicométricos em modelos preditivos de *machine learning*, a análise social torna-se essencial para equilibrar explicação e previsão (Shmueli, 2010), garantindo que as pesquisas possam utilizar

algoritmos que maximizem a precisão, mas também preservem o significado teórico, identifiquem vieses e assegurem o uso ético dos dados.

Os resultados deste trabalho instigam a reflexão sobre a importância de os analistas sociais, especialmente os psicólogos, se apropriarem de técnicas de análise usadas por outras áreas para analisarem temáticas historicamente trabalhadas pela psicologia. É importante refletir sobre essas metodologias, as informações que geram, as suas aplicações e implicações sociais, para um posicionamento ético e político dos analistas sociais e dos psicólogos sobre essas questões. Essas reflexões apontam para a necessidade de pesquisas futuras que permitam compreender e pensar sobre questões como: a) a formação do analista social e do psicólogo e o caráter multidisciplinar do uso de metodologias computacionais para pesquisas; b) a mudança do paradigma de pesquisa explicativo para o paradigma de pesquisa preditivo a partir de modelos computacionais e de aprendizagem de máquina; c) o compromisso ético e social da análise social e da psicologia no contexto de *Big Data*.

A formulação de métricas para dados secundários que correspondam a fenômenos psicológicos ou a elaboração de algoritmos preditivos exigem conhecimentos além daqueles vistos em sala de aula tradicionais nas áreas de ciências sociais e humanas. Se esses conhecimentos ainda não são exigência para a atuação profissional e produção científica na área, este cenário pode estar a transformar-se face às inovações tecnológicas recentes. Integrar métodos computacionais exige formação interdisciplinar, hoje em dia rara em psicologia e ciências sociais, gerando lacunas de competência técnica e teórica que dificultam a validação e a interpretação ética de algoritmos.

Há que discutir como fazer um uso ético de dados públicos disponíveis na internet. As práticas atuais de pesquisa com seres humanos exigem um termo de consentimento livre e esclarecido para participantes, mas há pouco debate sobre como proceder com dados que já foram produzidos e podem ser acessados em massa com o uso de tecnologias computacionais. Na pesquisa em psicologia e análise social, cabe às categorias profissionais avançar para esse debate com base nos princípios éticos tais como os estabelecidos no Código de Ética Profissional do Psicólogo (Conselho Federal de Psicologia, 2005); pela American Psychological Association (APA, 2017), e pelas diretrizes da International Sociological Association (ISA, 2019), entre outras, que sublinham a necessidade do respeito pela dignidade e autonomia dos sujeitos, e defendem a justiça social e a proteção de comunidades vulneráveis na pesquisa com dados digitais.

Como limitações deste trabalho, que apontam para uma agenda de pesquisa futura, assinala-se que não se realizou uma análise detalhada dos procedimentos metodológicos de cada pesquisa, e que, tendo em vista que cada

artigo utilizou metodologias com passos e amostras diferentes, essa descrição acrescentaria à compreensão sobre como a personalidade tem sido estudada no contexto do *Big Data*. Outra limitação: o trabalho analisa pesquisas publicadas em meados dos anos 2010/2020, período inaugural da análise psicométrica no contexto do *Big Data* nas redes sociais, e estudos futuros podem beneficiar de atualizações, incluindo estudos contemporâneos. Finalmente, em investigações futuras poder-se-ão explorar técnicas automatizadas de processamento de texto que permitiriam escalabilidade e redução de vieses na codificação.

Como consideração final, reforça-se que o mapeamento sistemático das manifestações digitais dos Cinco Grandes Fatores que aqui se realizou demonstrou a viabilidade técnica de unir psicomетria clássica e análise de *Big Data*, ao mesmo tempo que identificou também os riscos éticos e políticos que emergem dessa prática. Destaca-se, assim, a necessidade de que psicólogos e analistas sociais incorporem métodos computacionais com rigor teórico e responsabilidade ética. Essas descobertas apontam para um futuro em que a psicologia social poderá combinar predição e explicação, enformando intervenções mais eficazes em marketing, políticas públicas e saúde mental, ao mesmo tempo que defenderá textos regulatórios e práticas de governança de dados capazes de proteger a dignidade e a autonomia dos indivíduos em ambientes digitais.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGHAEI, S., BLACKWELL, A., KOSINSKI, M., STILLWELL, D. (2015), "Personality and intrinsic motivational factors in end-user programming". In *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Raleigh, NC, IEEE, 29-36.
- AGUIAR, M. A. F. (2005), *Psicologia Aplicada à Administração: Uma Abordagem Multidisciplinar*, São Paulo, Editora Saraiva.
- ALLPORT, G. W. (1927), "Concepts of trait and personality". *Psychological Bulletin*, 24 (5), pp. 284-293.
- ALLPORT, G. W. (1937), *Personality: A Psychological Interpretation*, Oxford, Holt.
- ALLPORT, G. W., ODBERT, H. S. (1936), "Trait-names: a psycho-lexical study". *Psychological Monographs*, 47 (1), pp. 1-171.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (2017), *Ethical Principles of Psychologists and Code of Conduct*. <https://www.apa.org/ethics/code/>.
- ANDREWS, A. L. (2018, 12 de abril). "The science behind Cambridge Analytica: does psychological profiling work?". Retirado de: <https://www.gsb.stanford.edu/>.
- BACHRACH, Y., KOSINSKI, M., GRAEPEL, T., KOHLI, P., STILLWELL, D. (2012). "Personality and patterns of Facebook usage". In *Proceedings of the 3rd Annual Web Science Conference ACM*, Nova Iorque, NI, ACM, pp. 24-32.
- BACK, M. D., STOPFER, J. M., VAZIRE, S., GADDIS, S., SCHMUKLE, S. C., EGLOFF, B., GOSLING, S. D. (2010), "Facebook profiles reflect actual personality, not self-idealization". *Psychological Science*, 21 (3), pp. 372-374.
- BARDIN, L. (1977), *Análise de Conteúdo*, Lisboa, Edições 70.
- BARROCAL, A. (2018, 19 de outubro), "As pistas do método 'Cambridge Analytica' na campanha de Bolsonaro". Retirado de: www.cartacapital.com.br.
- BOCK, A. M. B., ROSA, E. Z., AMARAL, M. M., FERREIRA, M. R., GONÇALVES, M. D. G. M. (2022), "O compromisso social da Psicologia e a possibilidade de uma profissão abrangente". *Psicologia: Ciência e Profissão*, 42, e262989.
- BRASIL (2018), Lei n.º 13.709, de 14 de agosto de 2018, dispõe sobre a proteção de dados pessoais e altera a Lei n.º 12.965, de 23 de abril de 2014 (Marco Civil da Internet).
- BRÜNING, C. (2015), *O Papel da Psicologia para a Administração*. Tese de doutoramento, Curitiba, Programa de Pós-Graduação em Administração, Universidade Federal do Paraná.
- CANTADOR, I., FERNANDEZ-TOBIAS, I., BELLOGIN, A., KOSINSKI, M., STILLWELL, D. (2013), "Relating personality types with user preferences in multiple entertainment domains". In *Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE)*, Boston, MA.
- CARVALHO, N. M. de, SILVA, C. M. da, NATIVIDADE, J. C. (2023), "Avaliação da personalidade na era do Big Data: uma revisão sistemática". *Cadernos de Psicologia*, 3 (3), pp. 1-23.
- CATTELL, R. B. (1943), "The description of personality: basic traits resolved into clusters". *The Journal of Abnormal and Social Psychology*, 38 (4), p. 476.
- CENTER FOR THE DIGITAL FUTURE (2017), *The 2017 Digital Future Report*. Los Angeles, CA, USC Annenberg.
- CHEN, E. E., WOJCIK, S. P. (2016), "A practical guide to big data research in psychology". *Psychological Methods*, 21, pp. 458-474.
- CHEN, P. C. L., ZHANG, C.-Y. (2014), "Data-intensive applications, challenges, techniques and technologies: a survey on Big Data". *Information Science*, 275, pp. 314-347.

- CONSELHO FEDERAL DE PSICOLOGIA (2005), Resolução CFP n.º 010/2005. *Código de Ética Profissional do Psicólogo*, XIII Plenário. Brasília, DF: CFP.
- COSTA, P.T., MCCRAE, R.R. (1992), "Normal personality assessment in clinical practice: the NEO Personality Inventory". *Psychological Assessment*, 4 (1), pp. 5-13.
- FARNADI, G., SITARAMAN, G., ROHANI, M., KOSINSKI, M., STILLWELL, D., MOENS, M.-F., DAVALOS, S., COCK, M.D. (2014), "How are you doing? Emotions and personality in Facebook". In *EMPIRE 2014: 2nd Workshop on Emotions and Personality in Personalized Services*, Boston, MA, pp. 45-56.
- FARNADI, G., SITARAMAN, G., SUSHMITA, S., CELLI, F., KOSINSKI, M., STILLWELL, D., DAVALOS, S., MOENS, M., COCK, M.D. (2016), "Computational personality recognition in social media". *User Modeling and User-Adapted Interaction*, 26, pp. 109-142.
- GANDOMI, A., HAIDER, M. (2015), "Beyond the hype: Big Data concepts, methods, and analytics". *International Journal of Information Management*, 35 (2), pp. 137-144.
- GOLDBERG, L.R. (1999). "A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models". In I. Mervielde, I.J. Deary, F. de Fruyt e F. Ostendorf (eds.), *Personality Psychology in Europe*, vol. 7, Tilburg: Tilburg University Press, pp. 7-28.
- GOLDBERG, L.R., JOHNSON, J.A., EBER, H.W., HOGAN, R., ASHTON, M.C., CLONINGER, C.R., GOUGH, H.G. (2006), "The international personality item pool and the future of public-domain personality measures". *Journal of Research in Personality*, 40, pp. 84-96.
- GOSLING S.D., GADDIS S., VAZIRE, S. (2007), "Personality impressions based on Facebook profiles". In *Proceedings of the International Conference on Weblogs and Social Media*; Boulder, CO e Menlo Park, CA: AAAI.
- GOSLING, S.D., KO, S.J., MANNARELLI, T., MORRIS, M.E. (2002), "A room with a cue: personality judgments based on offices and bedrooms". *Journal of Personality and Social Psychology*, 83 (3), pp. 379-398.
- GREENBERG, D.M., KOSINSKI, M., STILLWELL, D.J., MONTEIRO, B.L., LEVITIN, D.J., RENTFROW, P.J. (2016), "The song is you: preferences for musical attribute dimensions reflect personality". *Social Psychological and Personality Science*, 7 (6), pp. 597-605.
- GUZZO, R., FINK, A., KING, E., TONIDANDEL, S., LANDIS, R. (2015). "Big Data recommendations for industrial-organizational psychology". *Industrial and Organizational Psychology*, 8 (4), pp. 491-508.
- HABIB, F., ALI, Z., AZAM, A., KAMRAN, K., PASHA, F.M. (2024), "Navigating pathways to automated personality prediction: a comparative study of small and medium language models". *Frontiers in Big Data*, 7, 1387325.
- HAGGER-JOHNSON, G., EGAN, V., STILLWELL, D. (2011), "Are social networking profiles reliable indicators of sensational interests?". *Journal of Research in Personality*, 45 (1), pp. 71-76.
- HARLOW, L.L., OSWALD, F.L. (2016), "Big data in psychology: introduction to the special issue". *Psychological Methods*, 21 (4), pp. 447-457.
- HERN, A. (2018, 6 de maio), "Cambridge Analytica: how did it turn clicks into votes?". Retirado de www.theguardian.com.
- HUTZ, C.S., NUNES, C.H., SILVEIRA, A.D., SERRA, J., ANTON, M., WIECZOREK, L.S. (1998), "O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores". *Psicologia: Reflexão e Crítica*, 11 (2), pp. 395-411.
- IHSAN, Z., FURNHAM, A. (2018), "The new technologies in personality assessment: a review". *Consulting Psychology Journal: Practice and Research*, 70 (2), pp. 147-166.

- INTERNATIONAL SOCIOLOGICAL ASSOCIATION (2019), "ISA Code of Ethics". <https://www.isa-sociology.org/en/about-isa/isa-international-code-of-ethics>.
- JOHN, O. P., NAUMANN, L. P., SOTO, C. J. (2008), "Paradigm shift to the integrative Big Five trait taxonomy: history, measurement, and conceptual issues". In O. P. John, R. W. Robins e L. A. Pervin (eds.), *Handbook of Personality: Theory and Research*. Nova Iorque, NY, Guilford Press, pp. 114-158.
- KAPLAN, R. M., RILEY, W. T., MABRY, P. L. (2014), "News from the NIH: leveraging big data in the behavioral sciences". *Translational Behavioral Medicine*, 4 (3), pp. 229-231.
- KAUSHAL, V., PATWARDHAN, M. (2018), "Emerging trends in personality identification using online social networks – A literature survey". *TKDD*, 12 (2), pp. 15-30.
- KERN, M. L., EICHSTAEDT, J. C., SCHWARTZ, A. H., DZIURZYNSKI, L., UNGAR, L. H., STILLWELL, D. J., KOSINSKI, M., RAMONES, S. M., SELIGMAN, M. E. P. (2014), "The online social self: an open vocabulary approach to personality". *Assessment*, 21 (2), pp. 158-169.
- KIM, G.-H., TRIMI, S., CHUNG, J.-H. (2014), "Big-Data applications in the government sector". *Communications of the ACM*, 57, pp. 78-85.
- KOSINSKI, M., BACHRACH, Y., KOHLI, P., STILLWELL, D., GRAEPEL, T. (2014), "Manifestations of user personality in website choice and behaviour on online social networks". *Machine Learning*, 95 (3), pp. 357-380.
- KOSINSKI, M., BEHREND, T. (2017), "Editorial overview: Big Data in the behavioral sciences". *Current Opinion in Behavioral Sciences*, 18, pp. IV-VI.
- KOSINSKI, M., MATZ, S. C., GOSLING, S. D., POPOV, V., STILLWELL, D. (2015), "Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines". *American Psychologist*, 70 (6), pp. 543-556.
- KOSINSKI, M., STILLWELL, D., KOHLI, P., BACHRACH, Y., GRAEPEL, T. (2012). "Personality and website choice". *ACM Web Science Conference*, pp. 251-254.
- LAI, K., LEE, W. X., CHEN, H., RONGJUN, Y. (2017), "Research on web search behavior: how online query data inform social psychology". *Cyberpsychology, Behavior, and Social Networking*, 20 (10), pp. 596-602.
- LAMBIOTTE, R., KOSINSKI, M. (2014). "Tracking the digital footprints of personality". *Proceedings of the IEEE*, 102, pp. 1934-1939.
- LANDERS, R. N., BRUSSO, R. C., CAVANAUGH, K. J., COLLUMUS, A. B. (2016), "A primer on theory-driven web scraping: automatic extraction of big data from the Internet for use in psychological research". *Psychological Methods*, 21 (4), pp. 475-492.
- LIN, J. (2015), "On building better mousetraps and understanding the human condition: reflections on Big Data in the social sciences". *The Annals of the American Academy of Political and Social Science*, 659 (1), pp. 33-47.
- MACHADO, A. A. A. (2016), *Inferências da Personalidade a partir de Textos de Rede Social Utilizando Um Léxico Afetivo em Português Brasileiro*. Dissertação de mestrado, Sergipe, Universidade Federal de Sergipe.
- MAHALINGAM, V., STILLWELL, D., KOSINSKI, M., RUST, J., KOGAN, A. (2013), "Who can wait for the future? A personality perspective". *Social Psychological and Personality Science*, 5 (5), pp. 573-583.
- MARKOVIKJ, D., GIEVSKA, S., KOSINSKI, M., STILLWELL, D. (2013), "Mining Facebook data for predictive personality modeling". *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 7 (2), pp. 23-26.

- MATZ, S. C., KOSINSKI, M., NAVE, G., STILLWELL, D. J. (2017). "Psychological targeting as an effective approach to digital mass persuasion". *Proceedings of the National Academy of Sciences of the United States of America*, 114 (48), pp. 12714-12719.
- MAYER, J. R., MITCHELL, J. C. (2012), "Third-party web tracking: policy and technology. In 2012 IEEE Symposium on Security and Privacy, IEEE, Washington, DC, pp. 413-427.
- MAYER-SCHÖNBERGER, V., CUKIER, K. (2013), *Big Data: a revolution that will transform how we live, work, and think*, Boston, MA, Houghton Mifflin Harcourt.
- MCADAMS, D. P. (1997), "A conceptual history of personality psychology". In R. Hogan, J. A. Johnson e S. R. Briggs (eds.), *Handbook of Personality Psychology*, San Diego, CA, Academic Press, pp. 3-39.
- MESSENBERG, D., CAMARGOS, B. (2024), "Os propagandistas do ódio: o bolsonarismo-raiz em ação". *Análise Social*, 49 (253), e2318.
- MILLS, C. (2015), "Personality: technology, commodity and pathology". In I. Parker (ed.), *Handbook of Critical Psychology*, Londres, Routledge.
- MURPHY, S. C. (2017), "A hands-on guide to conducting psychological research on Twitter". *Social Psychological and Personality Science*, 8 (4), pp. 396-412.
- MYPERSONALITY PROJECT (maio, 2018). Retirado de www.myPersonality.org.
- NAVE, G., MINXHA, J., GREENBERG, D. M., KOSINSKI, M., STILLWELL, D., RENTFROW, J. (2018), "Musical preferences predict personality: evidence from active listening and Facebook likes". *Psychological Science*, 29 (7), pp. 1145-1158.
- NUNES, C. H. S. S., HUTZ, C. S. (2002). "O modelo dos Cinco Grandes fatores de personalidade". In R. Primi (org.), *Temas em Avaliação Psicológica*, Campinas, SP, IBAP, pp. 40-49.
- OPEN SCIENCE COLLABORATION (2015), "Estimating the reproducibility of psychological science". *Science*, 349 (6251), aac4716.
- OWEN, J., IMEL, Z. E. (2016), "Introduction to the special section "Big'er' Data": scaling up psychotherapy research in counseling psychology". *Journal of Counseling Psychology*, 63 (3), pp. 247-248.
- PARK, G., SCHWARTZ, H. A., EICHSTAEDT, J. C., KERN, M. L., KOSINSKI, M., STILLWELL, D. J., UGAR, L. H., SELIGMAN, M. E. P. (2015), "Automatic personality assessment through social media language". *Journal of Personality and Social Psychology*, 108 (6), pp. 934-952.
- PASSOS, M. F. D. (2014), *Elaboração e Validação de Escala de Diferencial Semântico para Avaliação de Personalidade*. Tese de doutoramento, Brasília, Universidade de Brasília.
- PENNEBAKER, J. W., FRANCIS, M. E., BOOTH, R. J. (2001), *Linguistic Inquiry and Word Count: LIWC2001*, Mahwah, NJ, Lawrence Erlbaum Associates.
- QUERCIA, D., CASAS, D. B., PESCE, J. P., STILLWELL, D., KOSINSKI, M., ALMEIDA, V. A., CROWCROFT, J. A. (2012a). "Facebook and privacy: the balancing act of personality, gender, and relationship currency". *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 6 (1), pp. 306-313.
- QUERCIA, D., KOSINSKI, M., STILLWELL, D., CROWCROFT, J. (2011), "Our Twitter profiles, our selves: predicting personality with Twitter". In *Proceedings of the Third International Conference on Privacy, Security, Risk and Trust (passat) and of the Third International Conference on Social Computing (socialcom)*, IEEE, Boston, MA, pp. 307-314.
- QUERCIA, D., LAMBIOTTE, R., STILLWELL, D., KOSINSKI, M., CROWCROFT, J. (2012b), "The personality of popular Facebook users". In *Proceedings of the Conference on Computer Supported Cooperative Work*, Nova Iorque, NI, ACM, pp. 955-964.

- REIS, T. (2018), *Personalidade e Big Data: Análise de Conteúdo dos Artigos Fundamentados na Base de Dados "myPersonality"*. Monografia, Curitiba, Universidade Federal do Paraná, curso de Psicologia.
- RENTFROW, P. J., GOSLING, S. D., JOKELA, M., STILLWELL, D. J., KOSINSKI, M., POTTER, J. (2013), "Divided we stand: three psychological regions of the United States and their political, economic, social, and health correlates". *Journal of Personality and Social Psychology*, 105, pp. 996-1012.
- RUTLEDGE, P. B. (2018, 21 de março), *How Cambridge Analytica Mined Data for Voter Influence*. Retirado de www.psychologytoday.com.
- SCHWARTZ, H. A., EICHSTAEDT, J. C., DZIURZYNSKI, L., KERN, M. L., BLANCO, E., KOSINSKI, M., STILLWELL, D., SELIGMAN, M. E., UNGAR, L. H. (2013), "Toward personality insights from language exploration in social media". Comunicação apresentada ao AAAI Spring Symposium: Analyzing Microtext, Palo Alto, CA.
- SCHWARTZ, H. A., EICHSTAEDT, J. C., KERN, M. L., DZIURZYNSKI, L., RAMONES, S. M., AGRAWAL, M., SHAH, A., KOSINSKI, M., STILLWELL, D., SELIGMAN, M. E., UNGAR, L. H. (2013), "Personality, gender, and age in the language of social media: the open-vocabulary approach". *PloS One*, 8 (9), e73791.
- SHMUELI, G. (2010). "To explain or to predict?". *Statistical Science*, 25 (3), pp. 289-310, <https://doi.org/10.1214/10-STS330>.
- SILVA, B. B. C. (2018), *Reconhecimento de Traços de Personalidade com Base em Textos*. Dissertação de mestrado, São Paulo, Universidade de São Paulo.
- SILVA, I. B., NAKANO, T. C. (2011), "Modelo dos cinco grandes fatores da personalidade: análise de pesquisas". *Avaliação Psicológica*, 10 (1), pp. 51-62.
- SIMCHON, A., GUNTUKU, S. C., SIMHON, R., UNGAR, L. H., HASSIN, R. R., GILEAD, M. (2020), "Political depression? A big-data, multimethod investigation of Americans' emotional response to the Trump presidency". *Journal of Experimental Psychology: General*, 149 (11), pp. 2154-2168.
- SIMÕES, M. J., AUGUSTO, F. R. (2019), "Expostos e duplamente vigiados". *Análise Social*, 54 (230), pp. 131-153.
- SOTO, C. J., JOHN, O. P. (2017), "The Next Big Five Inventory (BFI-2): developing and assessing a hierarchical model with 15 facets to enhance broad personality measurement". *Journal of Personality Assessment*, 99 (1), pp. 61-75.
- SRAVANTHI, K., REDDY, T. S. (2015), "Applications of Big Data in various fields". *International Journal of Computer Science and Information Technologies*, 6 (5), pp. 4629-4632.
- STEFANA, A., DAMIANI, S., GRANZIOL, U., PROVENZANI, U., SOLMI, M., YOUNGSTROM, E. A., FUSAR-POLI, P. (2025), "Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation". *Frontiers in Psychology*, 15, 1494261.
- STILLWELL, D., KOSINSKI, M. (2012), "*myPersonality project*: Example of successful utilization of online social networks for large-scale social research". Comunicação apresentada na International Conference on Mobile Systems (MobiSys).
- STILLWELL, D. J., Kosinski, M. (2015), *myPersonality Project website*.
- SUN, R. (ed.) (2008). *The Cambridge Handbook of Computational Psychology*, Nova Iorque, NY, Cambridge University Press, pp. 3-19.
- SWAN, K., SPEYER, R., SCHARITZER, M., FARNETI, D., BROWN, T., WOISARD, V., CORDIER, R. (2023), "Measuring what matters in healthcare: a practical guide to psychometric principles and instrument development". *Frontiers in Psychology*, 14, 1225850.

- TAVARES, M., (2010). *Avaliação Psicológica: Diretrizes na Regulamentação da Profissão*, Brasília, Conselho Federal de Psicologia, pp. 31-57.
- TECHAMERICA FOUNDATION'S FEDERAL BIG DATA COMMISSION (2012), *Demystifying big data: A Practical Guide to Transforming the Business of Government*.
- UNIÃO EUROPEIA (2016), Regulamento 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, "General Data Protection Regulation (GDPR)", <https://gdpr-info.eu/Regulamento>, *Jornal Oficial da União Europeia*, L 119, 4.5.2016, pp. 1-88.
- WAHEED, H., ANJUM, M., REHMAN, M., KHAWAJA, A. (2017), "Investigation of user behavior on social networking sites", *PLoS One*, 12 (2), p.e0169693.
- WATERFIELD, P., REVELL, T. (2018), *New Scientist*, 238 (3178), pp. 4-5
- YARKONI, T., WESTFALL, J. (2017), "Choosing prediction over explanation in psychology: lessons from machine learning". *Perspectives on Psychological Science*, 12 (6): 1100-1122.
- YOUYOU, W., STILLWELL, D., ANDREW SCHWARTZ, H., KOSINSKI, M. (2017). "Birds of a feather do flock together: Behavior-Based Personality-Assessment Method reveals personality similarity among couples and friends. *Psychological Science*, 28 (3), pp. 276-284.

Recebido a 07-06-2024. Aprovado para publicação a 24-11-2025.

REIS, T., BRUNING, C., MARQUES JUNIOR, K. (2026), "Traços de personalidade e Big Data: desafios éticos a partir de pesquisas realizadas com a base *myPersonality*". *Análise Social*, 259, LXI (2.º), e36295. <https://doi.org/10.31447/36295>.

Tharik Reis » thrkreis@gmail.com » Programa de Pós-Graduação em Psicologia, Universidade Federal do Paraná (UFPR) » Praça Santos Andrade, 50, 2.º andar, Centro — 80230-130 Curitiba, PR, Brasil » <https://orcid.org/0009-0007-3130-0187>.

Camila Bruning » camila.bruning@ufpr.br » Departamento de Psicologia, Programa de Pós-Graduação em Psicologia, Universidade Federal do Paraná (UFPR) » Praça Santos Andrade, 50, 2.º andar, Centro — 80230-130 Curitiba, PR, Brasil » <https://orcid.org/0000-0003-3869-3917>.

Karlo Marques Junior » karlomjunior@hotmail.com » Departamento de Economia, Programa de Pós-Graduação em Economia, Universidade Estadual de Ponta Grossa (UEPG) » Praça Santos Andrade, n.º 1, Centro, Campus Central, Bloco D, Sala 216 — 84010-919 Ponta Grossa, PR, Brasil » <https://orcid.org/0000-0003-2656-2637>.
