



# Data-Based User's Personality in Personalizing Smart Services

Izabella Krzeminska<sup>1,2</sup> 

<sup>1</sup> Poznan University of Economics and Business,  
al. Niepodległości 10, 61-875 Poznan, Poland

<sup>2</sup> R&D Labs, Orange Polska, Obrzeźna 7, Warszawa, Poland  
Izabella.Krzeminska@orange.com

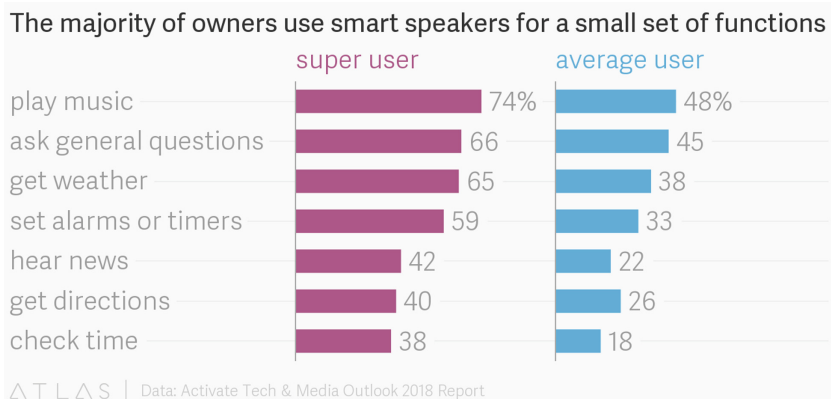
**Abstract.** Currently, a lot of attention and commitment are paid to improving customer experience with services. More and more proposed solutions are data-based, human-centred services. The purpose of this article is to present the assumptions and evidence in order to create a method that will allow for customization of the service functionalities (e.g. a smart home service with a virtual personal assistant) to the nature of the user (personality in the Big 5 model). The article will present the results of the preliminary research concerning users' differentiation of needs, the definition of research problems and the idea for a research scheme. The aim of the research is creating a model that classifies users based on their personality indicated from mobile phone data. What distinguishes the proposed solution from others, is that many types of different data are used (call logs, photos, applications, data from telephone usage history, etc.), which can be beneficial for assessment accuracy. Additionally, the proposed method allows for adapting the service to the user needs from the beginning of usage, without the necessity of collecting data about user activity.

**Keywords:** User profiling · Service personalizing · Data based services · Detecting personality based on digital data · Smart services · Sensitive services · Human-centred services

## 1 Introduction and Motivation

A continuous digital transformation has been taking place for several years. The habits and ways of carrying out tasks, as well as the general functioning of people in the world are gradually enriched by an increasing number of helpful digital tools. For digital advanced services (e.g. sensitive or smart home) it's crucial to reduce the barriers related to anxiety over artificial intelligence, such as virtual robots or assistants. On the other hand, good determinants of diversifying user needs are always sought. Profiling based on demographic characteristics (such as gender, age, place of residence, education, economic status) seems to be insufficient as needs' predictor. In the context of the ongoing digital transformation, traces of the digital activity of users are increasingly being used to collect information about the client, as well as for profiling or classifying [1,2].

The risk of service rejection and discontinuation of usage seems to be more important in the case of new, interactive and technologically advanced services based on artificial intelligence (intelligent services). This is due to individual differences and general concerns connected with new technologies. In addition, personalization seems to be beneficial before the service acquires data for profiling from activity history of user. An example to illustrate this problem can be the Amazon Alexa product, developed to encourage shopping by Amazon. In the report published in August 2018, we can find the claim “Of the people who did buy something using Alexa voice shopping, about 90% didn’t try it again” [3]. Rejecting a service after one use completely excludes analytics based on usage history. Moreover, analyzing the smart speaker usage report (see Fig. 1), it seems that users limit the use of the service to basic functionalities known from their own experience. As we can also see, the list of used functionalities is the same for super users and average users. Based on experience, it’s difficult for a user to find unknown functionalities available in the service, such as advanced shopping options. In this article [3], we can read that the representative of Amazon, who believes that the fact that so few people buy things through Alexa is a big challenge for the company, emphasizes also that the product will be able to be developed when the distinguishing features of individual customers are found. Also, recommendations based on analyzing behaviours in similar groups described by socio-demographic variables, or those based on segmentation will not allow for mapping any unused functionalities of the service. All these practices can be applied only if the services are actually used and are useless in the situation of leaving, discouragement or boredom of the user. One of the possible solutions is to have a basic usage profile already at the time of the service installation and to personalise the service from the first use according to the needs associated with the profile. The research program presented in the article has been developed simultaneously with the project of creating the service itself. Thanks to that the service, including the possibility of managing the



**Fig. 1.** The most popular question to smart speaker in group of super users (min. 3 times a day) and average service users. Source: [4].

functionalities of the service, is constantly adjusted to the results of the research on the classification of users. Taking the experience and effects of the work of other researchers, as well as the motivations resulting from the business need, presented above into account, the research program was created. In the article is proposed generic research framework for the detection of personality traits, which includes: (i) preparing the tool for data collection and creating a dedicated personality tool; (ii) a data processing stage for creating personality models and detecting the data required for personality detection; (iii) validation of usefulness of data-based automatically detected personality as a classifier of a service' user's needs.

## 2 Related Work

Personality has no single definition shared by all psychologists. In general, personality is a set of traits or characteristics, which are organised, relatively stable and which influence the cognition and behaviour of individuals [5]. Within the existing personality theories one of the most influential and numerous validated ones is the Big 5 model. The model has been developed mainly by Costa and McCrae since 1978 and in the 90s it was confirmed in a large number of empirical studies [6–9]. There also are number of studies showing strong relationship between personality and behaviour, life satisfaction and achievements and preferences e.g. [10,11].

In the past 10 years, many researchers and many companies have attempted to determine the user's personality based on different kinds of digital data. Most of the attempts concerned data from social media (Facebook, Twitter) [12–14] or other personal data like call logs [15] or mobile applications [16]. There is also evidence that this kind of data-based personality diagnosis is accurate [17,18]. Concerning methods used, models of indicating data-driven users' personality are mainly based on text analysis (e.g. tweets or FB posts) using statistical regression or simply r-Pearson correlation [13,19].

In 2013 [13] researchers proved that predicting personality based on telephone call logs data is possible. There is also a proven case of detecting personality based on Sociometric Badges after collecting for 6 months various kinds of data reflecting human behaviour [17]. One of the latest research in 2019 [21] concerned the detection of personality traits based on eye-tracking data as an alternative to the paper personality assessment tools. What is important, the purpose of this research was developing the clinical personality tool, which requires much better accuracy than tools for marketing purposes.

Tracking the digital footprint for detecting the users personality is broadly investigated by researchers using various kinds of large data sets like text, profile photo, music, film preferences based on the FB likes or relations (SNA) for example: [12,13]. Except for the analysis of the profile photo in social media, all other kinds of research were based on massive data, collected from the history of social service usage.

In 2017 authors in paper [20] confirm that the use of psychological targeting allows to influence the behaviour of the social media users by adapting the

message to the psychological needs of recipients defined by personality (Big 5). But at the beginning of using a new type of AI services, such as a virtual/voice assistant, data from the history of usage will not be available. In the context of even better regulations on the protection of users' privacy, we are slowly moving to a situation in which services will be limited to the use of own data, and data information systems will become more and more hermetic. Therefore, we are looking for methods that will allow services to detect user's personality based on even small amounts of data available at the time of service installation, e.g. from a mobile phone. In this paradigm, the assumption refers to determining the personality of each user of the smart home service, so the model should be based on the data available in each phone. Therefore, we have excluded such types of data which do not apply to everyone (e.g. social media or other specific services or applications).

### 3 Setting and Methods

#### 3.1 Personality as a Classification of the Need

For this research, the "Big Five" model is used to describe the personality of an individual. Big 5 model basically claims that there are five dimensional factors of personality.

- **Openness to experience** which describe tolerance for the new and the unknown.
- **Extraversion** describes tolerance for big quantity of stimulus.
- **Neuroticism** describes intolerance to stress.
- **Agreeableness** is about concentration on others' needs and willingness for co-operation.
- **Conscientiousness** is about intolerance to chaos and disorder.

The description of the factors as dimensional is very important. This means that these are not "yes or no" factors but all individuals are described by each factor to some extent. In other words, the dimensions describe the variability among people. So you can be a little bit or a lot more (or less) extraverted than average, but it wouldn't be accurate to say that you have no extraversion.

Survey conducted in 2018 (own research<sup>1</sup>), delivered clear evidence that personality dimensions are good enough for discriminating users needs and expectations (see Table 1). Moreover respondents confirm that the services based on

<sup>1</sup> The research was carried out in the first half of 2018, on 60 users of mobile phones, residents of Warsaw. In order to emphasize the diversity between people resulting from their personality, the study was conducted on a very homogeneous group in terms of demographics (the aim was to reduce confounding factors). The age has been limited to 20–29 years, separate groups for both sexes, technologically advanced group, using many functionalities of mobile devices. The study was a multi-stage: filling of the personality questionnaire (Big 5), monthly observation of behaviors in social profiles and in the use of telephone, in-depth structured interviews aimed at getting as much information as possible about behavior patterns. The results of these studies were used for creating behavioral metrics for each of Big 5 dimensions.

**Table 1.** Different attitudes and expectations connected with virtual assistant (VA) dependable on personality dimensions (Big 5) (source: own research)

Personality dimension	High level	Low level
Extraversion	Rejection of VA, but if adapted to the user expectations (no advertisements) can be considered	Reject based on privacy (need for high protection), VA in role of invisible friend
Openness for experience	The repeatability based on usage history is irritating but the idea of smart proactive VA searching all they need is attractive, they would be first to try it, their interest is changing and evolving	They expect to do something easier and faster - but still only when they need it (they initiate the action) - their interests is rather stable
Conscientiousness	The only group which like personalization, VA for them can be useful tool for controlling themselves, keep schedule and order	They are aware of their own problems so they want the functionalities that helps them avoid errors due to lack of systematic or discretionary decision-making etc.
Neuroticism	They appreciate the convenience of this solution (it limits forgetting, helps), but skeptical about data transfer and risk of “manipulation”	If they will have enough time for adaptation they can go with it (the value for them is an effectiveness in real life)
Agreeableness	They are afraid but if it became popular (recommended by friends) they will accept it	Their main motivation is a pragmatic needs (usefulness in everyday life)

artificial intelligence are currently not adapted to them and this is the primary reason for their rejection or dissatisfaction with available services. The short summary in tabular data set shows (Table 1) how diverse the attitudes and expectations of mobile phone users towards virtual assistants are. Differences concern mainly the function of a virtual assistant and are a simple derivative of diversity of needs. At the same time, it can be seen that attitudes polarize with respect to personality dimensions. For example, people who are open to experience expect non-standard content, have a high level of cognitive needs, and high need to explore (curiosity). In turn, people with low openness expect only a sense of comfort in a world that is well known to them (they like only what they know, they are afraid of unknown). For another dimension, individuals with a high level of conscientiousness use mainly the functionalities that help in the implementation of the need for control, which mainly manifests itself in the control of time and in scrupulous planning. On the other hand people with

a low level of conscientiousness, who accept life in chaos and disorder, need only very basic control functionalities and will never be interested in using advanced calendar or notebook functions.

Due to the specificity of research and linking it with electronic services, the decision of creating dedicated Big 5 tool was taken. Diagnostic tools for personality measurement are available on the market in the form of licensed commercial questionnaires (e.g. NEO Personality Inventory Revised) and NEO-FFI (NEO Five-Factor Inventory), created by Costa and McCrae (1992) [22,23]. These tools are used primary for individual diagnosis and they are under protection. Psychological diagnostic tools cannot be disseminated, because it weakens their diagnostic power, nor be modified. The procedure of using them is very strict and can be conducted only by authorized psychologists. Another reason was the need for integration of Big 5 tool with electronic applications. The whole psychometric Big 5 tool creation procedure was processed from the beginning. In this way creation of a dedicated tool became possible and easy. Due to business needs, personality dimensions (Extraversion, Agreeableness, Conscientiousness, Neuroticism (Emotional Stability) and Openness to experience, will be treated independently - as if we were creating 5 separate tools. The issue of orthogonality of factors will also be examined due to theoretical foundations and due to the problem of interference between factors in the situation of building a model based on data.

### 3.2 Objectives of the Research and Thesis

Having in mind the presented motivation for the research, the following research questions were defined:

- RQ1. How to create method for determining the user's personality based on the available data (from mobile phone)?
- RQ2. How to evaluate the reliability of the method of personality detecting?
- RQ3. How to evaluate the utility of developed method in real service/data?

The main goal of the dissertation is to propose effective and accurate methods of assessing the user's needs defined by the dimensions of his personality (e.g. BIG 5) based on the data available from the moment of service installation without the delay connected with collecting user's data log. The specific objectives of the work are:

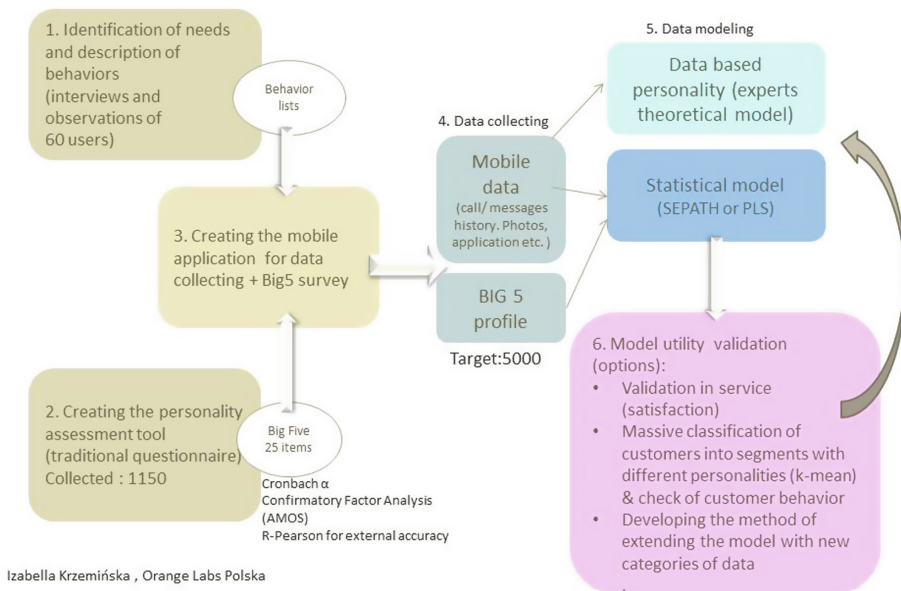
- Develop the method for determining user's personality based on the limited amount of data
- Develop reliability indicators for the method of personality determination based on the available amount of user data without having to collect usage history
- Develop the method for verifying a pre-designated personality based on dynamic data when using the service
- Assess the applied needs classifier by experiment based on laboratory measurement or measurement on a test group

- Assess reliability and accuracy of personality measurement (Big 5) based on user data in comparison to declarative data (personality questionnaire tool)

Hence the thesis was formulated as follows: **Automatic personality assessment (Big 5), based on user data of any interactive service, e.g. a virtual assistant, can be used as a classifier of user’s needs for the way of interaction and the set of service functionalities, which will significantly speed up profiling of the service and will increase satisfaction with using it.**

## 4 Research Scheme

The overall methodology chosen for conducting required research is Design Science [24]. The main paradigm of Design Science Research, proposed by Hevner, is answering research questions by creating innovative artifacts, and in this way enhancing existing knowledge and science base by new knowledge [25]. In design science paradigm the main goal is to build and evaluate an artifact. To verify the defined thesis and to accomplish the research objectives, the following research scheme is designed (see Fig. 2):



**Fig. 2.** Research scheme & methods

1. **Preliminary qualitative tests** carried out to identify needs and collect descriptions of discriminating behaviour.

2. **Creating the personality assessment tool**, which can be used for this specific research (low number of questions (25) and tool tested and created for online or mobile app usage). Standard psychometric procedures are applied.
3. **Developing the mobile application** dedicated for collecting data from mobile phone and Big 5 assessment. So far the data from the following user activities (sources) have been collected and analyzed on a mobile phone of a user:
  - Telco data – call logs statistics, text messages logs statistics etc. (collected from mobile phone logs)
  - Application data – applications on mobile phone, phone parameters, etc.
  - Photos – photos in user photo albums, etc.
  - Phone settings and statistics e.g. battery consumption, kind of security level, etc.
4. **Main research fieldwork** - data collecting required for creating the model - planned 5000 participants. This is the current point of the research.
5. **Creating the personality model** based on the mobile phone data, available at the moment of service installation (single drop without additional collecting of the activity logs). There is an idea to create at least 2 different models. The first will be theoretical model coherent with psychological Big 5 theory. The second will be pure statistical model based on SEPATH or PLS. From psychological point of view, it will be a very interesting and unique possibility of comparing these two differently created models.
6. **Validation stage** e.g. in interactive service which allows experimental manipulation based on personality adapted service. Or in laboratory simulation way comparing satisfaction metrics or checking the purchasing behaviour after a massive classification of the whole data base of mobile customers.

This is the plan of work stages. The presented user-oriented research program has been developed simultaneously with the project of creating the service itself. Thanks to that the service, including the possibility of managing the functionalities of the service, is constantly adjusted to the results of research on the classification of users. The research is now on stage 4 - collecting data for creating models. Stages 1, 2 and 3 have already been finished. The mobile application for data collection with implementing 25 questions tool for Big 5 assessment is ready.

The future works concern stages 5 and 6. The validation stage (6) is now rather a list of proposals and is not completely defined. The final shape will highly depend on the phase of maturity of the created parallel service, as well as on the availability of data on service platform. Stage 5 and stage 6 will run cyclically until satisfactory results are achieved. It will also be repeated if the data-set is expanded or the assumptions changed.

The planned comparison of the result based on behavioral metrics and personality test results can bring to very interesting conclusions. It can be said, that personality questionnaires reflect the image of how we see ourselves and not necessarily real behavior. If this image is biased (for various reasons, e.g. high need for social approval or cognitive defense mechanisms) the diagnosis will not be



accurate. Moreover, the answers in the personality questionnaires are sometimes based on subjective perspective, e.g. the assessment of the amount of time spent with friends and social gathering may not be adequate to the real place in the distribution for the population. So if we create a model of predicting the personality questionnaire results (declarative), we will repeat this error. Therefore, in the research scheme, we included the creation of a personality profile based on theory and actual reference to the results of the whole representation of population. So potentially it will be possible to verify the declared behavior with the actual ones. Of course another issue is the interpretation of this possible gap.

## 5 Conclusion and Discussion

The preliminary research suggests that personality, as a predictor of user's needs, can significantly facilitate profiling of services. The results, based on the observation of behaviors related to the use of a mobile phone and experience with services based on AI algorithms, confirm the theory of personality in the Big 5 approach. It seems that extending the range of diagnostic tools measuring users' natural needs is an obvious step in the development of intelligent and sensitive services. The most important advantage of the proposed solution is not being limited to one type of data and basing the model on the data available at the time installing the service. This perfectly complements the currently used profiling methods (based on data from user activity registers). Fitting the service to the user's needs from the moment of installing the service can be crucial in prevention against rejection of innovative, technologically advanced and unknown services, such as a virtual assistant in a smart/sensitive home.

An additional contribution to the current state of knowledge is that the target solution will function on different types of data (call history, texts, photos). As a result, the solution will be more versatile and easier to implement in other applications. The planned research aims to provide evidence of the effectiveness of using the user's personality determined from digital data and significantly increase the satisfaction of using these services. The validation phase still needs to be refined and it will be a big challenge to find reliable measures to assess both user satisfaction and the accuracy of the profile of personality detected. Moreover comparing the effectiveness of differently created model (theoretical and statistical) can be interesting evidence for further psychological research.

The research carried out in this scheme will probably make a significant contribution to the discussion on the reliability of personality diagnostics based on questionnaires, in which information on behavior is declarative. This has significant consequences for the subsequent stages of work and for assessment of reliability and accuracy of profile detecting methods.

The potential of a solution that can be developed towards a faster psychological diagnosis is also important. Using a similar research program, tools for diagnose other features such as addiction to the phone, neuroses, depression and burnout can be created. In turn, the method of implementing knowledge about the user's personality in the service is relatively easy to implement in other types

of activities such as creating applications and interfaces. It can be the first step for the development of automatically adaptable service interfaces. It is worth emphasizing that the presented solution is not an alternative to profiling solutions based on logs of activity. If the presented type of profiling proves to be effective, it can be used as a basis. In the future, IT systems should be designed in the way that will allow for validation of the initial model based on activity data. In the case of significant lack of coherence between initial profile and user choices, they will be able to update it.

## References

1. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. (PNAS)* **110**(15), 5802–5805 (2013)
2. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns of Facebook usage. In: *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 24–32. ACM (2012)
3. Anand, P.: The Reality Behind Voice Shopping Hype. <https://www.theinformation.com/articles/the-reality-behind-voice-shopping-hype?> Accessed 15 Apr 2018
4. Activate-tech-media-outlook-2018. <https://www.slideshare.net/ActivateInc/activate-tech-media-outlook-2018>. Accessed 15 Apr 2019
5. Strelau, J.: *Psychologia Temperamentu: Wydawnictwo Naukowe PWN* (2019)
6. Costa Jr., P.T., Mc Crae, R.R.: Four ways Five Factors are basic. *Personality Individ. Differ.* **12**(13), 653–665 (1992)
7. McCrae, R.R., Costa Jr., P.T.: Toward a new generation of personality theories: theoretical context for the Five-Factor Model. In: Wiggins, J.S. (ed.) *The Five-Factor Model of Personality: Theoretical Perspective*, pp. 51–87. Guilford, New York (1996)
8. De Raad, B.: Structural models of personality. In: Corr, W.P.J., Methews, G. (eds.) *The Cambridge Handbook of Personality Psychology*, pp. 127–147. Cambridge University Press, New York (2009)
9. Goldberg, L.R.: The development of markers for the Big-Five factor Structure. *J. Pers. Soc. Psychol.*, 99–110 (2016)
10. Barrick, M., Mount, M.: The Big Five Personality, dimensions and job performance: a meta-analysis. *Pers. Psychol.* **44**(1), 1–26 (1991)
11. Judge, T., Higgins, C., Thoresen, C., Barrick, M.: The Big Five Personality traits, general mental ability and career success across the life span. *Pers. Psychol.* **52**(3), 621–625 (1999)
12. Liu, L., Preotiuc-Pietro, D., et al.: Analyzing Personality Through Social Media Profile Picture Choice. *Association for the Advancement of Artificial Intelligence* (2016)
13. Kern, M.L., et al.: *The online social self an open vocabulary approach to personality*. University of Pennsylvania, University of Cambridge (2013)
14. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter profiles, our selves: predicting personality with Twitter. In: *IEEE SocialCom* (2011)
15. de Montjoye, Y.-E., Quoidbach, J., Robic, F., Pentland, A.: *Predicting Personality Using Novel Mobile Phone-Based Metrics*. MIT, Harvard University, Ecole Normale Supérieure de Lyon (2013)

16. Xu, R., Frey, R.M., Fleisch, E., Ilic, A.: Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. ETH Zurich (2016)
17. Kalimeri, K., Lepri, B., et al.: Going beyond traits: multimodal classification of personality states in the wild. MIT, ACM, 13 December 2013. ISBN 978-1-4503-2129-7
18. Back, M.D., et al.: Facebook profiles reflect actual personality not self-idealization. *Psychol. Sci.* **21**(3), 372–374 (2010). <https://doi.org/10.1177/0956797609360756>
19. Goldbeck, J., et al.: Predicting personality from Twitter. In: IEEE International Conference on Privacy, Security, Risk, and Trust (2011)
20. Matz, S., Kosinski, M., Nave, G., Stillwell, D.: Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl. Acad. Sci.* **114** (2017). 201710966. <https://doi.org/10.1073/pnas.1710966114>
21. Berkovsky, Sh., et al.: Detecting personality traits using eye-tracking data. In: CHI Conference (2019). <https://doi.org/10.1145/3290605.3300451>
22. Costa, P.T., McCrae, R.R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Psychological Assessment Resources, Odessa, FL (1992)
23. Zawadzki, B., Strelau, J., Szczepaniak, P., Sliwiska, M.: Inwentarz Osobowości NEO-FFI Costy i McCrae. Adaptacja polska. Professional Manual. Warsaw, PTP (1998)
24. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. *MIS Q.* **28**(1), 75–105 (2004)
25. Hevner, A., Chatterjee, S.: Design Research in Information Systems: Theory and Practice. Springer, Heidelberg (2010). <https://doi.org/10.1007/978-1-4419-5653-8>