International Journal of Data Science and Analytics (2026) 22:39
https://doi.org/10.1007/s41060-026-01026-8

REGULAR PAPER

A data-driven analysis to predict energetic intelligence

Luis Gonzaga Baca Ruiz[1] · David Criado-Ramón[2] · María José Serrano-Fernández[3] · Elena Pérez-Moreiras[4] · María del Carmen Pegalajar Jiménez[2]

Received: 26 November 2024 / Accepted: 7 January 2026

Abstract

Energetic Intelligence is a newly defined construct recently validated that offers a thorough understanding of human intelligence by integrating, emotional, spiritual, and physical dimensions. This study applies a data-driven approach to predict Energetic Intelligence. Our research goal lies in the application of machine learning techniques to model and predict Energetic Intelligence using data commonly gathered in organizational contexts. We collected responses through structured surveys and employed a range of supervised learning algorithms to build predictive models. Model performance was evaluated using standard metrics, with the best results reaching an R^2 of 0.73 through optimized and simplified models, which is a promising outcome for a psychologically grounded prediction task. Key predictors included variables such as Flow, Flourishing, and Emotional Vitality, which consistently emerged as relevant features in model training. These findings demonstrate the potential of machine learning to support psychological research and offer practical tools for the assessment and development of Energetic Intelligence in applied settings. Our work points out the value of integrating AI methodologies with psychological theory to enable data-driven insights into human potential and well-being.

Keywords Machine learning · Data mining · Energetic intelligence · Psychology

Abbreviations

AB	AdaBoost
AI	Artificial Intelligence
BR	Bagging Regressor
CB	CatBoost
DT	Decision Tree
EN	Elastic Net
EngI	Enegetic Intelligence
ET	Extra Trees
GB	Gradient Boosting
LGBM	LightGBM
LR	Linear Regression
MLP	Multi-Layer Perceptron
R^2	R-squared
RF	Random Forest
RT	Regression Tree
SR	Stacking Regressor
XGB	XGBoost

✉ María del Carmen Pegalajar Jiménez
mcarmen@decsai.ugr.es

Luis Gonzaga Baca Ruiz
bacaruiz@ugr.es

David Criado-Ramón
dcriado@ugr.es

María José Serrano-Fernández
mariajose.serrano@urv.cat

Elena Pérez-Moreiras
eperez-moreiras@gruporh.com

1 Department of Software Engineering, University of Granada, 18014 Granada, Andalucía, Spain

2 Department of Computer Science and Artificial Intelligence, University of Granada, 18014 Granada, Andalucía, Spain

3 Department of Psychology, Rovira I Virgili University, 43002 Tarragona, Spain

4 RH Asesores Improving S.L, Instituto Para El Desarrollo del Talento Natural y El Nuevo Liderazgo, 28221 Madrid, Spain

1 Introduction

Human intelligence is a key aspect of knowledge that has intrigued people throughout history because it is the most distinctive feature of humans. While traditional approaches have explored its philosophical and psychological dimensions, recent advances in data science offer powerful new tools to analyse and predict cognitive and behavioural patterns with greater precision. The integration of data-driven methods allows researchers to move beyond theoretical constructs and toward measurable, actionable perceptions. In recent years, the rise of data science has revolutionized the way we study and interpret intelligence, transforming it into a domain that can be better quantified, modelled, and optimized. By leveraging large datasets, machine learning algorithms, and computational models, researchers can reveal patterns in cognitive performance, simulate aspects of intelligent behaviour, and develop predictive tools with practical applications. This data-driven approach is critical in addressing modern challenges such as climate change, global conflict, and the ethical deployment of Artificial Intelligence (AI). Understanding intelligence through the lens of data science enables not just theoretical insight but also actionable solutions that enhance human adaptability, mental health, and societal resilience.

Energetic Intelligence [1] (hereinafter, EngI) is inspired by various existing intelligence models, highlighting a fundamental element, energy, which, although it has been present in the study of this construct since the beginnings of its scientific study, had not been sufficiently highlighted. EngI also proposes an integrative framework, incorporating linguistic, thought, emotional, spiritual, bodily, and movement aspects of human behaviour, offering a perspective that connects and expands previous approaches. Below is an analysis of its evolution as a construct, as well as theories and research that support this perspective.

The study of intelligence has evolved significantly throughout history. The earliest signs date back 2.5 million years to a group of Australopithecus [2]. The author notes the emergence of several species of the genus Homo, which were characterised by larger brain and tool use, resulting in great evolutionary flexibility. These species controlled fire, cared for each other, and were successful hunters, but their social structure was very rigid, leading to their extinction. Only the African branch (sapiens) has survived through us and has inhabited the Earth for over 20,000 years. This ability to think is the first evidence of intelligence in our ancestors.

In ancient times, 2500 years ago, philosophers like Aristotle reflected on the nature of intelligence, viewing it as a unique human ability. Aristotle defines the *psyche* as the «first entelechy of an organised natural body», the principle of life and the realisation of potential. This Aristotelian interpretation of the soul forms the theoretical basis of our EngI, since psychology should address not only behaviour but also the essence of the human being [1]. Plato, considered the origin of rational psychology, divides the soul into three parts: rational (immortal, in the brain), irascible (mortal, in the thorax), and concupiscent (mortal, in the abdomen) [3].

Etymologically, intelligence comes from the Latin words "inter", meaning "between", and "legere", meaning "to choose". Therefore, intelligence relates to the human ability to "choose between", to exercise one's freedom and free will, which is the most notable trait that distinguishes humans from other living beings [1].

With the emergence of psychology as a science during the nineteenth century, intelligence became one of the most studied constructs in psychology, and researchers sought ways to define and measure it. Among them, many emphasise its energetic nature. Johann F. Herbart states that ideas are like atoms with energy and consciousness and can attract or repel other ideas based on their compatibility [4]. William James highlights the importance of the total available energy for initiating mental and moral operations, noting how each person recognises the difference between days when energy flow is high and when it is low [5]. Today, data science offers powerful tools to quantify these once-abstract notions. And it enables researchers to analyse cognitive patterns and mental processes through empirical data and computational techniques.

In 1904, one of the key figures in the construct, Spearman, defined the g factor of General Intelligence as mental energy [6, 7]. In 1905, Binet and Simon developed the first intelligence test [8], pioneering a design for testing that has been refined to this day. David Wechsler introduced 'purpose' as part of the components of intelligence, defining it as the set of activities to act with purpose, think rationally, and effectively deal with the environment [9, 10], creating two instruments for its measurement, one for adults and another for children.

At the end of the twentieth century, there is a significant expansion in the definition of intelligence. In addition to the previously mentioned theories that focus on cognitive and energetic aspects, new theories highlight various skills and abilities, as well as emotional and spiritual aspects of being human. This view of intelligence aligns with trends in data science, where emotional intelligence is increasingly valued. It helps data scientists interpret human behaviour and collaborate effectively in order to enhance both analysis and impact [11].

Anderson [12] developed the ACT-R model, which stands for Adaptive Control of Thought-Rational. In this model, W is a constant divided by all the elements a person focuses on while performing a task. The "attentional energy" parameter, or resource activation (W), affects how accurately and quickly even simple tasks are completed.

In the same decade, many authors have explored a broader view of intelligence. Howard Gardner proposed the Theory of Multiple Intelligences [13]. He argued that intelligent behaviour does not come from a single quality of the mind but from different types of intelligences based on separate reserves of mental energy. Each of these reserves allows individuals to solve problems and create valuable products in one or more cultural contexts. Gardner identifies eight intelligences: (1) linguistic, (2) logical-mathematical, (3) spatial, (4) Body and Movement, (5) musical, (6) interpersonal, (7) intrapersonal, and 8) naturalistic.

In 1985, Reuven Bar-On developed the EQ-i: Bar-On Emotional Quotient Inventory as part of his doctoral thesis to measure social and emotional competencies. Mayer, Salovey, and Caruso created the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), which evaluates four components of Emotional Intelligence: the ability to perceive, understand, manage, and use emotions for adaptive and creative thinking. Later, in 1997, Danah Zohar introduced the concept of Spiritual Intelligence with Marshall [14]. They define it as the ability to solve problems, find meaning, and express values [15]. Zohar sees Spiritual Intelligence as an aspect of intelligence that focuses on conscious meaning and purpose, beyond traditional IQ and various notions of Emotional Intelligence. In this context, Robert A. Emmons defines Spiritual Intelligence as the ability to use spiritual information adaptively to solve everyday problems and achieve goals. He states that Spiritual Intelligence makes individuals aware of transcendent reality and allows them to seek unity and reach their full potential [16]. EngI includes Zohar's Spiritual Intelligence as one of the five factors that comprise it.

Additionally, Gardner added the ninth intelligence, Existential Intelligence, to his list in 1998 and 1999. He defines Existential Intelligence as the cognitive capacity to raise and ponder "big questions" (i.e., queries about love, about evil, about life and death) about the nature and quality of existence".

Robert J. Sternberg developed the Triarchic Theory of Human Intelligence, which states that intelligent behaviour comes from a balance of analytical, creative, and practical skills. His Triarchic Model of Intelligence [17, 18] focuses on how intelligence is directed (cognitive style) rather than how much intelligence a person has (cognitive ability) [19]. In 1986, he noted that some definitions of intelligence distinguish between two types: capacities, like speed and mental energy, and dispositions, such as being self-critical. Along with Yang [20], he reviewed Chinese philosophical views on intelligence and concluded that the Confucian perspective emphasizes benevolence and doing what is right. An intelligent person puts in effort to learn, enjoys learning, and persists in lifelong learning with enthusiasm. A key feature of this theory is adaptability, both within the individual and in the sociocultural context [21]. Sternberg emphasised that the same information processing components are involved in all types of skills; what differs is how they are applied. He highlights a fourth type of thinking, Wise Thinking, which relates to a person's ability to contribute to the common good, both short-term and long-term. In 2002, he suggested it might be time to expand the conventional idea of intelligence to include not just memory and analytical skills, but also creative and practical skills. In 2018, he pointed out that a higher IQ is not always relevant for solving today's world problems. He presented a model of limited resources as a complement to the Theory of Successful Intelligence, emphasising the importance of social (1), especially creative (2), and practical (3) skills based on wisdom, compared to analytical skills (4). He defines Successful Intelligence as the ability to formulate, implement, evaluate, and, if necessary, reformulate life plans. This definition includes creative, analytical, and practical thinking, along with wisdom [18]. For Sternberg, intelligence is: (1) the ability to achieve one's goals in life, given each person's sociocultural context; (2) leveraging strengths and correcting or compensating for weaknesses; (3) adapting, shaping, and selecting environments; and (4) using a combination of analytical, creative, and practical skills, later adding identifying skills. All these contributions reflect a strong movement that advocates for the inclusion of uniquely human aspects related to the exercise of freedom to choose, the transitive and spiritual dimensions of intelligence, as well as the meaning of life and the purpose of existence, which Wechsler had already addressed in the mid-twentieth century but which, until the beginning of this twenty-first century, had not been sufficiently considered.

Recently, Pérez-Moreiras [1] integrated all these soft conceptions of intelligence (complementary to IQ) and formulated EngI as the ability to identify the energies within and outside oneself, distinguish between them, and use that information to achieve personal and collective goals aligned with one's Purpose of Life. It differs from immediate mood, as it represents a dynamic ability that can be improved rather than a temporary emotional state. The model (see Fig. 1) is based on the universal principle that everything that exists is information, energy and matter. This model is the result of an optimal experience [22] to serve as a graphic guide for the development and study of the construct.

EngI builds upon principles that have been refined in recent years into the The theory of Unified Energetic Intelligence in development. The theoretical basis rests on the idea that everything has a purpose in itself [23] and that energy and matter are interdependent, transforming without being created or destroyed [24]. Information is also a physical reality, not merely an abstraction.

Human beings are understood as informed energy in motion with purpose. They can perceive, manage and channel energy, matter and information towards personal and collective goals, using free will. Human life is systemic and energetic, manifested through linguistic, emotional, bodily and spiritual domains. Environmental energy also influences behaviour. Levels of consciousness range from unconscious to expanded, with pathological altered states as exceptions. Developing awareness of these energies and their effects

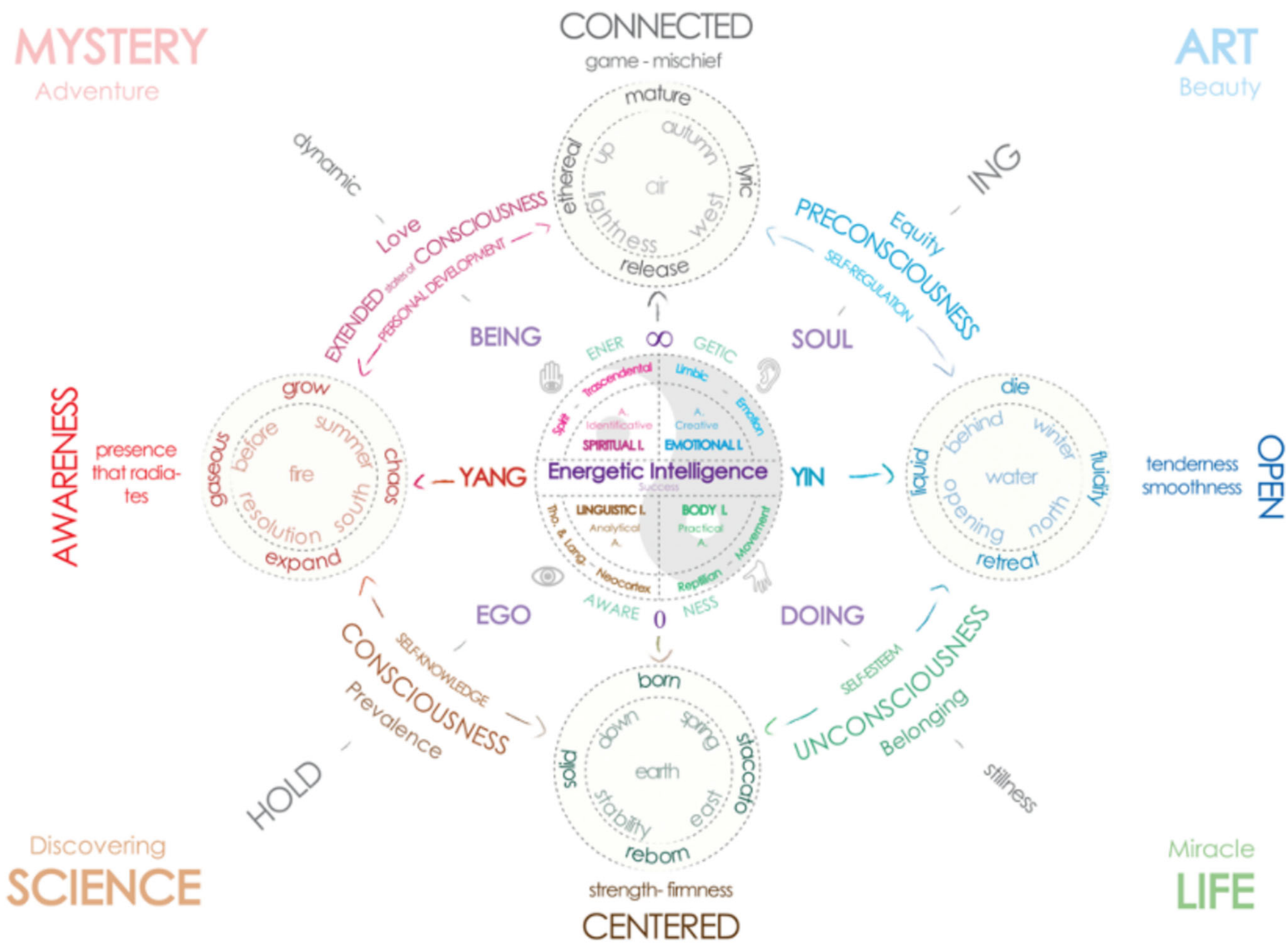

Fig. 1 Representation of the expanded model which includes Coach State, Five Rhythms, Family Systemic Configurations, Dispositions to Movement and Model Skills [1]

constitutes Energetic Intelligence, which enhances human capacity and purpose [1].

To psychometrically operationalise the construct, five initial factors were chosen (see Fig. 2): four intelligences (Body and Movement, Emotional, Linguistic and Transitive-Spiritual) and a fifth factor called Energetic Awareness. The analysis results revealed a correlation between EngI, Flow, Flourishing, Self-Efficacy, Self-Esteem, and most personality traits according to the Big Five Model. While certain components align with established constructs, EngI integrates these within a unified framework centred on the conscious management of informed energy. This integration extends beyond traditional psychological domains by incorporating the unitary energetic dimension of human functioning and all existence, the result of three fundamental components: information, energy and matter.

Psychometric validation has identified five second-order factors that structure EngI. The first, Body and Movement Intelligence, refers to the ability to recognise and constructively use bodily information and movement, with studies showing strong links between energy, physical activity, well-being and self-esteem. The second, Emotional Intelligence, concerns the recognition and use of emotions for individual and collective goals, as emotion mobilises vital energy for action and creativity while influencing mental health and resilience. The third, Linguistic Intelligence, highlights the role of language and thought in shaping reality, solving problems and fostering well-being, in line with research on language as central to cognition, identity and framing processes. The fourth, Transitive and Spiritual Intelligence, reflects the ability to find meaning and life purpose while transcending the self to serve the greater good, associated with vitality, optimism and commitment. Finally, Energy Awareness refers to awareness of one's energetic dimension, the capacity to regulate it and to create clean energetic environments, with evidence that high levels of energy foster productivity, creativity and optimal experience.

Furthermore, this proposed intelligence find resonance with, and also offer expansions to, established psychological theories. For instance, the Body and Movement and

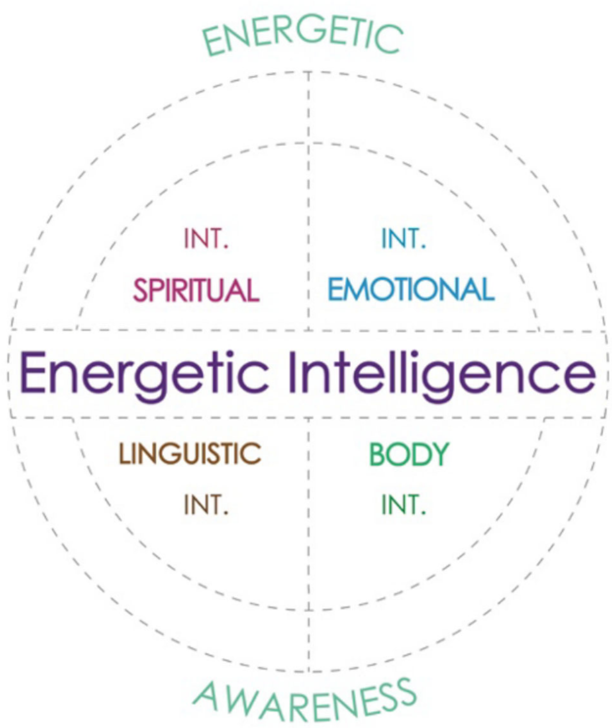

Fig. 2 Model of the validated EngI [1]

Linguistic Intelligences described here align with Howard Gardner's theory of multiple intelligences [13], but they introduce subtle emphases: Body and Movement Intelligence highlights the conscious processing of information from one's own and others' bodies, and Linguistic Intelligence underscores a transformative capacity to "shape realities". Similarly, Emotional Intelligence incorporates core components of Daniel Goleman's model of Emotional Intelligence, such as self-awareness and empathy in recognizing and managing emotions in oneself and others [25], while distinctively integrating an energetic dimension. Aspects of Transitive-Spiritual Intelligence and Energetic Awareness also touch upon Gardner's intrapersonal and existential intelligences [13] and can be seen as both contributing to and being enhanced by Emotional Intelligence's components of self-awareness and self-regulation [25].

Moreover, Transitive-Spiritual Intelligence builds upon concepts akin to Danah Zohar's Spiritual Intelligence, which emphasizes meaning, vision, and values [26]. However, it distinguishes itself by focusing on the dynamic fluidity and movement between ego-based and transcendent states of consciousness as a key operational aspect of applying spiritual insight. Energetic Awareness, while more specific in its focus on subtle energy fields than the broader scope of Spiritual Intelligence, could offer an experiential pathway to some of Spiritual Intelligence's underlying themes, such as interconnectedness or the "intelligence of the soul".

EngI is a unified integrative framework focused on the conscious management of informed energy. This integration transcends traditional psychological domains by incorporating the unitary energetic dimension of human functioning and existence as a result of information, energy, and matter. In today's complex world, being energetically intelligent becomes important, vital, and a priority in all social areas (biological, ecological, economic, educational, business, political, family, etc.). EngI provides a holistic view connected to life purpose [9, 10] and the energetic nature of human beings, which, as shown earlier, was already hinted at by early intelligence researchers in the early twentieth century [6, 7] but had not been clearly formulated in an evident, additional, and integrative way until now. This additive view of the soft factors of intelligence brings order and coherence to human behaviour, which is inherently unified. The only place where peace must be ensured to preserve it globally is within each human being. In light of the significant challenges humanity is experiencing, it is especially important to help people: (1) identify, become more aware, and be more competent in using the energy within and around them; (2) learn to regulate it; (3) choose how and where to invest it; (4) experience flow states in adversity; and (5) identify their life purpose that gives meaning to their existence, regardless of the external circumstances they face.

Despite psychologists' efforts to enhance their questionnaires and approaches, implementing these improvements in real life remains challenging due to the complexity of human behaviour, participant engagement or even time constraints. In this context, AI serves as a valuable tool for psychologists, helping to address these challenges effectively [27]. Utilizing advanced algorithms and data analysis features, AI can enhance this process. Additionally, AI can assist in analysing large datasets, identifying patterns and insights that may not be readily apparent through traditional methods [28]. Besides, this technology can enhance participant engagement by offering interactive and user-friendly interfaces, which may result in more precise and significant data collection. Furthermore, through data analysis and machine learning methods, we can expedite various processes and enhance efficiency by learning from user features. By identifying specific patterns in user behaviour, these methods enable us to skip certain steps in assessments or interventions when users exhibit particular characteristics [29]. This predictive ability enables a more customised model, as it can foresee needs and reactions based on specified characteristics [30]. In doing so, we can enhance the experience for both psychologists and participants.

In this work, we aim to analyse the data collected and develop an AI model that can predict an individual's EngI based on factors commonly measured in companies. This prediction is made without prior measurements of EngI or its components, providing information about the individual's current energetic state and helping them achieve greater well-being in their life. While EngI is a relatively new and lesser-known psychological variable, our study takes a Computer Science approach to investigate its predictability through data analysis. By optimising algorithms for extracting related variables from psychological questionnaires, we aim to contribute to the understanding of this emerging trait. Our focus on the computational aspects of EngI prediction is driven by the potential of data-driven insights to enhance psychological research. This approach allows us to explore the capabilities of machine learning and genetic algorithms in predicting EngI, with the goal of providing a foundation for further research in this area. The final goal of this paper is an artificial intelligence tool whose use must be guided by ethical considerations to prevent potential misuse. Therefore, the responsible interpretation of these assessments requires careful oversight. They should not serve as definitive judgments but as supportive tools that complement human expertise.

Before introducing the methodology, we would like to make a brief clarification. In the field of psychology, many variables are latent, meaning they cannot be directly observed, and are therefore referred to as constructs. As a result, these constructs are typically measured indirectly using questionnaires or other assessment tools. This important distinction allows us to distinguish between more general psychological variables and specific personality factors.

The following sections of this paper are arranged as indicated follows. In Sect. 2, we provide an in-depth description of the proposed methodology. It discusses the specific instruments, participants, machine learning techniques, measures, and metrics utilised in the research. Subsequently, Sect. 3 details the experiments, their setup, and procedures. The experimental outcomes are presented in Sect. 4. Concluding the research, Sect. 5 defines the drawn conclusions and future work.

2 Methodology

The current section introduces the proposed methodology. It goes through data analysis to eventually develop forecasting models to predict our target variable, Total EngI. It can be seen in the following flowchart (see Fig. 3), which outlines the key steps of the process.

The initial phase of our methodology involved data collection. Raw data were gathered through questionnaires administered to a diverse group of individual workers who are international Spanish speakers. This approach was chosen to ensure a comprehensive dataset reflecting a variety of perspectives and experiences. Following collection, the data went through a preprocessing and transformation stage. This involved a three-step process: first, cleaning the data to address any inconsistencies or errors; second, transforming variables as needed to suit analytical requirements; and third, formatting the data to ensure it was in a suitable structure for subsequent analysis.

Once the data were prepared, exploratory data analysis was conducted. This step involved performing statistical analyses and examining correlations within the dataset to uncover initial patterns, relationships, and insights. The findings from the exploratory analysis served as a guide for the subsequent model development phase.

The core of our methodology then moved into an iterative model building phase, designed to develop forecasting models for our target variable, Total EngI. This phase was characterized by the exploration of several distinct forecasting strategies. For each selected strategy, a model was developed and then evaluated. The process finishes after a certain number of iterations. If not, the process could loop back to refine the current strategy for a new iteration of model building and evaluation. We followed several strategies within this framework. A direct strategy, which focused on straightforward predictions. A stratified approach was employed, aiming to improve model performance. An oversampled approach was also implemented independently to address any class imbalances and assess its impact on effective model training. Furthermore, a personal variables approach was adopted, in which we incorporate additional personal variables to check potential enhancements to predictive capabilities. Finally, a genetic algorithm for feature selection was applied. Our aim was to develop a simplified model that utilised fewer variables while still maintaining good predictive performance.

2.1 Instruments

The study's instruments are detailed below. The following scales and subscales serve as predictor variables. The selected variables used in the prediction model were derived from completely different questionnaires, which assess distinct aspects of human psychology. Our analysis confirmed that the variables do not have a strong pairwise relationship, which supports the use of prediction models in this context. The lack of relationship between the variables implies that the models are not simply simplifying or replicating existing relationships, but rather identifying complex patterns that provide non-trivial insights.

The General Self-Efficacy Scale [31], now in Spanish [32], uses a four-point response scale, where 1 represents strong disagreement and indicates a complete lack of agreement, and 4 signifies total agreement. The belief in one's ability to handle daily life stressors is quantified as General Self-Efficacy. The internal consistency coefficient, α, was 0.87.

Some authors [33] have adapted the Psychological Well-Being Scale [34, 35] for use with Spanish-speaking workers.

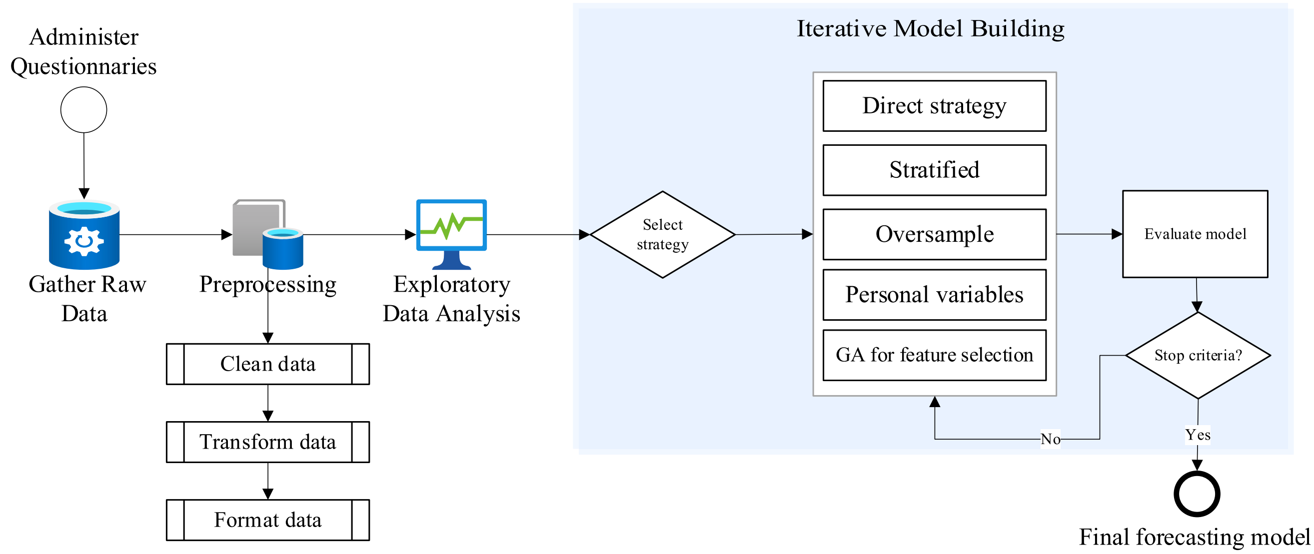

Fig. 3 Methodology overview

The scale comprises 5 items ($\alpha = 0.87$), with 7 response alternatives ranging from Strongly disagree (1) to Strongly agree (7). The internal consistency of this measure is strong, as indicated by a Cronbach's alpha coefficient of 0.88. This scale measures an individual's psychological well-being in relation to flourishing or personal growth, encompassing feelings of happiness and well-being. The correlation between these results and other scales of psychological well-being and feelings is strong, as suggested in [35].

The Short Dispositional Flow Scale [36] was translated into Spanish [37] using a sample of Spanish athletes. With a α coefficient of 0.80 for the nine-item scale. This scale employs a Likert-type response format with five options ranging from 'I never experience these sensations' to 'I always experience these sensations.

The adapted Spanish version of the Rosenberg Self-Esteem Scale [38] consists of 10 items evaluated using a four-point Likert scale, ranging from 1 (disagree strongly) to 4 (agree strongly). Five items were written positively and five negatively. With an α of 0.86. For example, 'I can perform tasks like most people'.

According to [39], a Spanish version of the Satisfaction with Life Scale was adopted. This scale, with a coefficient of 0.84, consists of 5 items. The response format utilizes a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

Pérez-Moreiras's EngI Inventory (ENII-33) evaluates individuals' capacity to differentiate and use the energy they perceive within and outside themselves to reach personal and group objectives according Life Purpose [1]. This survey consists of 33 questions, each presented with a five-point Likert scale ranging from strongly disagree to strongly agree.

Cronbach's α values for each factor are (f1) Bodily & Movement Intelligence ($\alpha = 0.84$); (f2) Emotional Intelligence ($\alpha = 0.86$); (f3) Linguistic Intelligence (α = 0.86); (f4) Transitive-Spiritual Intelligence ($\alpha = 0.90$); and (f5) Energetic Awareness ($\alpha = 0.91$).

The Overall Personality Assessment Scale [40] is a 42-item test based on the Big Five personality model, using a five-point scale from 1 (completely disagree) to 5 (completely agree). It includes five factors: Extraversion (7 items, $\alpha = 0.86$), Emotional Stability (7 items, $\alpha = 0.86$), Conscientiousness (7 items, $\alpha = 0.77$), Agreeableness (8 items, $\alpha = 0.71$), and Openness to Experience (8 items, $\alpha = 0.81$).

The Personal and Organizational Quality Assessment Revised 4 Scale [41, 42] assesses work environment quality through four subscales: Emotional Vitality, Emotional Stress, Organizational Stress, and Physical Stress. It uses a Likert scale from 1 (never) to 7 (always), with α values ranging from 0.76 to 0.92 across the subscales.

The Spanish version of the Trait Meta-Mood Scale [43] is a widely used Emotional Intelligence scale in psychological and educational research. It consists of three dimensions: Emotional Attention, Clarity, and Repair. Each containing 8 items. Reliability analysis shows strong internal consistency of 0.89 for Emotional Attention, 0.84 for Emotional Clarity, and 0.83 for Emotional Repair.

This research dataset comprises data on multiple topics, encompassing 12 personal factors, 16 occupation-related factors, and 26 psychological traits, all collected without initial filtering.

2.2 Participants and procedure

This study included 2203 Spanish-speaking workers were divided into 22.8% men and 77.2% women. Among the variables collected, we can find the gender, age, marital status and employment status. Educational background and professional information reported academic level, job title, tenure in the company, tenure in the current position, tenure in the profession and professional group. Organisational context variables captured the company sector, approximate number of employees and the geographical scope of the company's operations. Additionally, participants provided information on the economic situation of their company over the past year, as well as subjective assessments of their work environment, including ratings of their department and company climate, leadership quality of their supervisor, team spirit and collaboration, attitude of colleagues, self-assessed job performance, and the degree of alignment with both colleagues and supervisors.

To provide a brief overview of the participants, the average age was 46.09 years (with a $\sigma = 10.54$). 50% were married, 34% were single, 15% were divorced or separated, and 1% were widowed in terms of marital status. Besides, 46% possessed university education, while 2% had primary, 22% had secondary, and 28% had master's or doctoral studies. For this research, non-probabilistic sampling, also known as accidental random sampling, was used. The workers participated anonymously and voluntarily, without their personal identification being recorded. The objectives of the study were clearly explained to them. The study was carried out in accordance with the Declaration of Helsinki and the protocol followed the guidelines of the Ethics Committee of the participating university, which gave its approval.

Psychological and social research often encounters difficulties when dealing with large sample sizes. Despite having a larger sample size than many other studies in the field, our research has its limitations, which will be discussed in the results section.

The data for this study were collected through online questionnaires administered to consenting working individuals. Participation was voluntary and confidential, and participants were encouraged to respond honestly. Anonymity was guaranteed throughout the process. Both the participants and the organizations they worked for were selected based on accessibility, using a non-probability sampling approach, specifically, random sampling [44].

We would like to mention that participants received detailed information about the study before participation. Participants were informed that the survey contained questions regarding their personal, professional and emotional states and that they should answer honestly by selecting the option that best described them without overthinking, emphasising spontaneity. The process took approximately 20 to 30 min, after which participants received a provisional report summarising their responses and the concepts studied. Also they were assured that all data collection complied with applicable data protection regulations in Spain and the USA, and that electronic questionnaires were administered via an encrypted platform ensuring participant anonymity throughout the study.

2.3 Machine learning algorithms

In this section, we provide a brief introduction to the models. We acknowledge that a deeper overview of all 15 models would be extensive. Rather than providing detailed descriptions, we recognize that these models are well-established and widely recognized in the field of machine learning. For readers seeking more in-depth information, we recommend consulting relevant references, as the models are already well-documented. We have intentionally avoided providing excessive detail, as it would not add significant value to the discussion, and instead, we focus on the key aspects that are relevant to our context. Having said that, we implemented several algorithms, each with its own advantages and weaknesses. We primarily implemented four categories of algorithms: linear models, tree-based models, other ensemble methods, and neural networks.

In the category of linear models [45], we used Linear Regression (LR), which aimed to establish a relationship between the dependent variable and one or more independent variables. Ridge Regression (Ridge) was then applied, which introduced a penalty term to the loss function to prevent overfitting by adjusting the coefficients. Lasso Regression (Lasso) also focused on regularisation but had the added benefit of performing variable selection by forcing some coefficients to be exactly zero. Lastly, we employed Elastic Net (EN), which combined the strengths of both Ridge and Lasso. This approach allows for a balance between the two regularization techniques [46].

A set of tree-based models was considered in this study [47]. We started with Regression Trees (RT), which provided a simple but effective way to model relationships in many scenarios by splitting the data into subsets based on feature values. We then added Random Forest (RF), which is an ensemble method that built multiple decision trees and averaged their predictions to enhance accuracy and control overfitting. Extra Trees (ET) also contributed to this category by creating a collection of trees but with a more randomised approach to feature selection [48]. We explored Gradient Boosting (GB), which sequentially builds trees, each correcting the errors of the previous ones, leading to a strong predictive model. XGBoost (XGB) took this a step further by optimising the boosting process for speed and performance. We also implemented LightGBM (LGBM), which is known for its efficiency with large datasets [49]. We included CatBoost (CB) in our set of models because it typically performs exceptionally well with categorical features [50]. In addition, we included the Bagging Regressor (BR), which combined the predictions of multiple models to improve stability and accuracy.

We also included other ensemble methods, namely AdaBoost (AB), which focuses on adjusting the weights of misclassified instances to improve the model's performance iteratively [51]. And we also utilised the Stacking Regressor (SR), which combined the predictions of multiple models, specifically XGB, LGBM, and GB, to create a more robust final prediction [52]. We chose to include these three in our SR due to their proven effectiveness and complementary strengths. We considered other alternatives, however, we found that they did not offer significant improvements over the selected models.

Lastly, we explored Neural Networks, specifically the Multi-Layer Perceptron (MLP). This model consisted of multiple layers of interconnected nodes, allowing it to learn complex patterns in the data through backpropagation.

Therefore the 15 models implemented are: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Regression Trees, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, CatBoost, Bagging Regressor, AdaBoost, and a Stacking Regressor and Multi-Layer Perceptron.

2.4 Genetic algorithm

A GA is an optimization technique inspired by the principles of natural selection and genetics [53]. In our study, we employed a GA for feature selection, which helped us identify the most relevant features from our dataset. This approach mimicked the process of natural selection, where the algorithm generated a population of potential solutions and then iteratively improved them over several generations. Each solution represented a subset of features, and we evaluated their performance based on how well they contributed to the overall model accuracy. By selecting the best-performing subsets, the algorithm gradually eliminated less useful features while retaining those that provided the most value. The GA's goal is to build a simplified model that does not suffer in performance and allows us to reduce our feature set while enhancing the model's effectiveness. By reducing noise and focusing on the most informative variables, we ensured that the final model remained robust and accurate.

In this binary problem, we are dealing with a situation where each variable can either be included in the model training or excluded from it. This is represented by a binary value, where a value of 1 indicates that the variable will be used, while a value of 0 signifies that it will not. Consequently, each individual in our population is represented as a binary vector composed of these 1 s and 0 s, reflecting the selection of variables for training the model.

The process begins with the initialisation of a population, where we create a set of random individuals. Following this, we apply various genetic operators to evolve the population over successive generations.

Our GA employs mutation as a critical mechanism. The random changes introduced to the population help prevent premature convergence on suboptimal solutions and maintain genetic diversity. Based on a defined mutation rate, we randomly change the values of each feature in an individual during our mutation process.

Crossover is another essential operator that combines the genetic information of two parental individuals to produce offspring. We implemented two methods for crossover: uniform crossover and one-point crossover [54]. In uniform crossover, we create a mask of random binary values that determines which features are inherited from each parent. This results in two children that share characteristics from both parents. In one-point crossover, we select a random point along the length of the parent individuals and exchange the segments beyond that point to create two new offspring. Note that both crossover operators were tested, but no significant difference was found. Therefore, the results presented in this study were obtained using the uniform crossover method.

An individual's fitness is determined through the difference in their results and those of a test set. The fitness score indicates the degree to which the selected variables enhance the model's performance. The goal is to determine the optimal mix of features for training to boost the model's predictive capabilities.

2.5 Measures

Measuring the performance of our model requires the application of appropriate metrics. We use R-squared (R^2) as the chosen metric. Although additional evaluation metrics such as Root Mean Square Error and Mean Absolute Error were calculated during model development, they were not included in the final results, as they provided no significant additional vision beyond what was already captured by the R^2 values. Given the number of algorithms tested and the volume of resulting data, including these metrics would have unnecessarily lengthened the results section without enhancing interpretability. R^2 was selected as the primary evaluation metric because it directly reflects the proportion of variance explained by each model. Also, it is a particularly relevant measure in psychological and behavioural research where understanding explained variance is more meaningful than absolute error magnitudes.

R^2 can range from $-\infty$ to 1, where 1 indicates a perfect fit, and values less than 0 suggest that the model performs worse than a basic mean-based model. The R^2 metric is calculated using the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{1}$$

where $\overline{y}$ represents the mean of the actual values:

$$\overline{y} = \frac{1}{n} - \sum_{i=1}^{n} y_i \tag{2}$$

We also measured the skewness of the variables to assess the asymmetry of their distributions. This statistic reveals the degree of the data's deviation from normal distribution, aiding in the identification of outliers and the comprehension of data patterns. We will make use of this metric to better understand the nature of the data. It can be calculated as follows, being σ the standard deviation:

$$S = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{y_i - \overline{y}}{\sigma}\right)^3 \tag{3}$$

Kurtosis is a statistical measure that describes the shape of a probability distribution, particularly highlighting the tails and peak. It can be used to reveal information about the presence of outliers and the degree of "tailedness" compared to a normal distribution. Its calculation is shown in the following equation.

$$Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{y_i - \overline{y}}{\sigma}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{4}$$

3 Experiments

For those who are not concerned with our study's experimental details, they can skip this section. The process is described in detail by describing our experimental parameters. Thus, this section explains the specifics of our experiments, including each test and its evaluation parameters. To fully grasp our methodology, the following lines cover all essential factors of each model. Our model's performance was evaluated using five-fold cross-validation. In this way, we reduce overfitting, improving the model's ability to estimate its generalisation.

To optimise the performance of our models, we performed a grid search for parameter tuning. The best results were obtained by systematically exploring the optimal settings for each parameter. All the tested parameters are listed below.

For LR, we focused on the intercept parameter and tested both the inclusion (True) and exclusion (False) of the intercept.

The Ridge model included parameters such as the regularisation strength with values of 0.1, 1.0, 10.0, and 100.0, solver options including 'auto', singular value decomposition, Cholesky decomposition, least squares, and gradient-based optimisers, along with the maximum number of iterations was set to 1,000, 2000, and 3000.

In the case of Lasso Regression, we examined regularisation strengths with values of 0.01, 0.001, 0.0001, 0.1, 1.0, 10.0, and 100.0, maximum iterations of 1000, 2000, and 3000, selection methods of cyclic and random, and whether to reuse the previous solution (warm start) as True or False.

The DT involved parameters such as the criterion for splitting set to mean squared error, splitter options of best and random, maximum depth values of None, 10, 20, and 30, minimum number of samples required to split set to 2, 5, and 10, minimum samples per leaf with values of 1, 2, and 4, and the number of features to consider for the best split set to 2, 4, 8, and 16.

For RF, we included the number of trees in the forest ranging from 2 to 256, the splitting criterion set to mean squared error, maximum depth options of None, 10, 20, and 30, minimum number of samples to split set to 2, 5, and 10, minimum samples per leaf values of 1, 2, and 4, the number of features to consider for the best split set to 2, 4, 8, and 16, and whether to bootstrap samples as True or False.

The MLP model featured different hidden layer sizes, activation functions including rectified linear unit, tanh, and identity, solver options of 'adam', stochastic gradient descent, and limited-memory Broyden–Fletcher–Goldfarb–Shanno, regularisation strength values ranging from 0.0001 to 1.0, learning rate strategies of constant, inverse scaling, and adaptive, initial learning rate values of 0.001 and 0.01, maximum iterations set to 200, 300, and 500, early stopping enabled as True, and validation fraction set to 0.1.

For GB, we tested the number of boosting stages set to 50, 100, and 200, learning rate values of 0.01, 0.1, and 0.2, maximum depth options of 3, 5, and 7, and the minimum number of samples to split and the minimum number of samples per leaf both set to 2, 5, and 10.

The AB model includes the number of boosting stages set to 50, 100, and 200, along with learning rates of 0.01, 0.1, and 1.0.

For XGB, we examined the number of estimators set to 50, 100, and 200, learning rate values of 0.01, 0.1, and 0.2, maximum depth options of 3, 5, and 7, minimum child weight values of 1, 2, and 3, and subsample options of 0.6, 0.8, and 1.0.

In the case of BR, we included the number of estimators ranging from 10 to 200, maximum samples of 0.5 and 1.0, maximum features set to 0.5 and 1.0, and whether to bootstrap samples as True or False.

For ET, we tested the number of trees set to 50, 100, and 200, the number of features to consider for the best split with values of the square root and $\log_2$ of the feature count, maximum depth values of None, 10, 20, and 30, minimum samples to split set to 2, 5, and 10, and minimum samples per leaf values of 1, 2, and 4.

The LGBM model included parameters such as the number of boosting iterations set to 50, 100, and 200, and the learning rate values tested at 0.01, 0.1, and 0.2. The number of leaves was tested with values of 31, 50, and 100. We also evaluated the maximum depth of the trees, with options of -1 (no limit), 10, and 20. The minimum number of data points required in a leaf was set to 20, 50, and 100. In addition, the feature fraction, which determines the proportion of features to use in each boosting iteration, was set to 0.8 and 1.0.

The CB model has the following parameters. The amount of boosting iterations was tested with values of 50, 100, and 200. We explored different learning rates of 0.01, 0.1, and 0.2, along with tree depth options of 4, 6, and 8. Moreover, we included the L2 regularisation strength with values of 1, 3, and 5. We also tested the bagging temperature with values of 0, 0.5, and 1, and the number of splits used to make categorical feature combinations with options of 32, 50, and 100.

For EN, we evaluated the regularisation strength with values of 0.01, 0.1, 1.0, 10.0, and 100.0, and the mixing ratio between Lasso and Ridge penalties with values ranging from 0.00001 to 1.0. The maximum number of iterations was 1000, 2000, and 3000, while the tolerance for optimisation was tested at 0.0001, 0.001, and 0.01. Additionally, we tested whether to fit the intercept in the model, with options of True or False.

Finally, the SR model combined predictions from multiple models, including XGB, LGBM, and GB, using LR as the final estimator. We repeated the tests carried out previously for these models.

Once we have completed all but the last set of experiments, we fixed the best parameters obtained from our grid search to launch the GA. The parameters tested are a population size of 100, a total of 100 generations, a mutation rate of 0.1, and two types of crossover methods: uniform and two-points. Additionally, we evaluated the impact of elitism, which determines whether the best individuals from one generation are transferred to the next, with options for both true and false.

To minimise bias from the search space and reduce overfitting to specific patterns in the training data, we applied repeated cross-validation and executed all experiments ten times and used the mean results.

Table 1 Summary statistics of Total EngI. P_k stands for the k-th percentile

Statistic	Value
Total	2190
Mean	0.6579
S. deviation	0.1736
Min	0
P_{25}	0.5530
P_{50}	0.6742
P_{75}	0.7803
Max	1
Skewness	− 0.5359
Kurtosis	0.2583

Due to the limited space available in this paper, only the most relevant experiments are discussed below. In doing so, we hope to maintain a clear focus on our key findings.

4 Results and discussion

In this section, we present the results of our study. It begins with a brief exploratory analysis that provides some understanding into the data. After this initial examination, we introduce a detailed exploration of the forecasting experiments conducted.

As mentioned in previous sections, the dataset used in this study included information about employees, encompassing several aspects such as demographics, job details, perceptions of the work environment, personality traits, and emotional intelligence scores. It comprised a total of 2,190 rows and 52 columns, which featured both personal and psychological variables. The main focus of this study was the target variable known as «Total EngI». This variable has no missing values, with a mean value of approximately 0.658 and a standard deviation of 0.174. It ranges from 0 to 1, and the median is around 0.674, see Table 1.

The last two rows of the table provided information as to the asymmetry and tail characteristics of the data distribution for the variable in question. The analysis revealed a negative skewness of -0.53, indicating that the distribution was left-skewed. This suggested that higher scores were more prevalent. Additionally, the Kurtosis value of 0.25 indicated that it exhibited positive kurtosis. This suggested that the distribution had heavier tails and a greater likelihood of extreme values, which is characteristic of what is known as a leptokurtic distribution. Overall, these statistics point out the nature of the data distribution and emphasise the tendency for

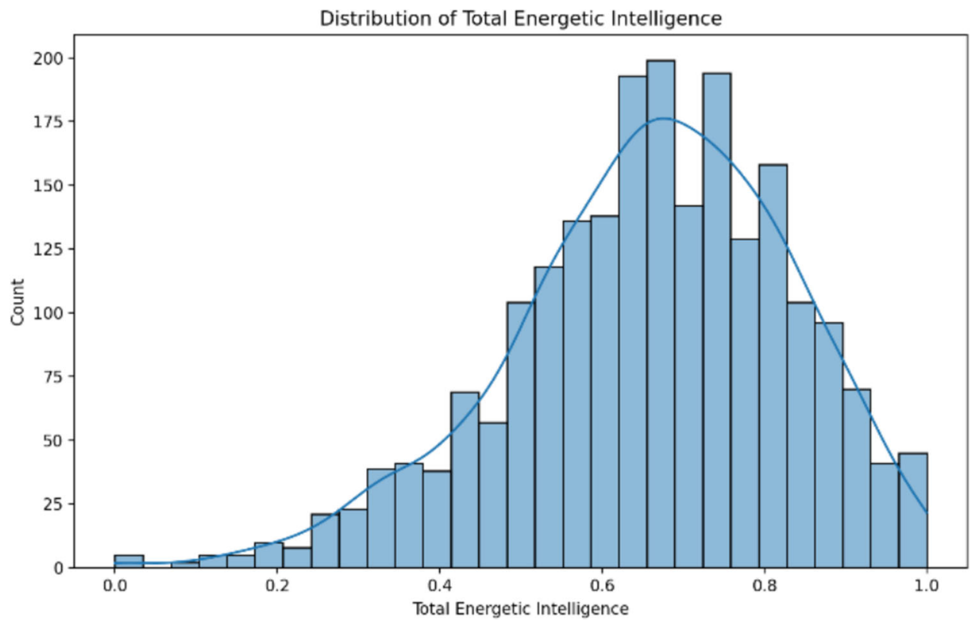

Fig. 4 Bar chart showing the frequency distribution of Total EngI scores among participants

higher scores to dominate while still allowing for the presence of outliers on the lower end. This can be visualized in Fig. 4.

In examining the relationships within the dataset, we identified the top three psychological variables that were most strongly correlated with our target variable. First, Flow, with a correlation of 0.686, stands out as a significant contributor. Emotional Vitality followed closely with a correlation of 0.669; and Flourishing, which had a correlation of 0.650.

In contrast, the personal variables showed much weaker correlations. It implies that they had less impact on our target variable. For instance, Position Tenure had a correlation of 0.002, while Company Tenure and Professional Experience followed with correlations of 0.004 and 0.007, respectively. Marital Status, with a correlation of 0.014, also demonstrated a minimal relationship with EngI. This remarkable difference highlighted that psychological factors are far more influential than personal circumstances. This might suggest that focusing on emotional and psychological well-being could be key to enhancing overall EngI capabilities.

Having concluded the exploratory analysis, we now turn our attention to the results of the forecasting experiments.

In the first experiment, we aim to compare the effectiveness of two distinct methods for splitting our datasets into training and testing subsets: a random procedure and a stratified approach. The stratified method is particularly noteworthy as it allows for a more subtle distribution of classes within the datasets. To explore this further, we tested three variations of the stratified approach concerning singleton classes, in order words, those classes that appear only once in the dataset. Specifically, we examined the impact of placing these classes in the test set, the training set, or in both sets. It is expected that including singleton classes in both the training and testing datasets will provoke overfitting, but we added this option to see how much the models could actually improve.

Table 2 summarizes the performance of various models across different scenarios. Note that in order to consolidate the results into a single table, there are several experiments that are not included since they did not provide extra information. The ET model stood out in scenarios where singleton classes were included in both training and testing sets. Across the remaining configurations, Ridge, LR, and MLP consistently delivered strong performance, while RT showed markedly lower results overall. Ensemble models such as RF, XGB, and CB generally performed well but did not always outperform simpler models like Ridge or LR. Particularly, performance for most models (excluding RT) fell within a similar range of 0.590 to 0.690.

Although models tended to perform better when singleton classes were included in both datasets, results still improved when singleton data were isolated in the test set, supporting the decision to adopt stratified splitting in subsequent analyses. This strategy ensured better generalization by testing the models on unseen data, which ultimately led to slightly higher R^2 values. On the contrary, when singleton data were present in training, performance declined, suggesting overfitting risks.

Regarding the oversampling experiment, we evaluated various thresholds between 10 and 60. Overall, Lasso showed the most stable and highest average performance with 0.651, especially at mid-range thresholds. MLP excelled only at the lowest threshold, 10, but showed a significant drop at higher levels, likely due to overfitting. As oversampling increased,

Table 2 Results using the R^2 metric. Best values in each column are bolded, while best values in each row are underlined

Model	Simple	Singleton in test	Singleton in train	Psycho Vars	Stratified	Oversample	Gender	Personal vars
LR	0.650	0.679	0.651	**0.674**	0.679	0.666	**0.672**	0.671
Ridge	**0.652**	**0.681**	**0.652**	0.673	**0.681**	0.665	0.671	0.670
Lasso	0.650	0.679	0.640	0.673	0.679	0.665	0.672	0.670
RT	0.345	0.508	0.418	0.436	0.508	0.451	0.529	0.398
RF	0.637	0.651	0.639	0.653	0.651	0.652	0.666	0.664
MLP	0.651	0.679	0.651	0.670	0.679	**0.667**	0.666	0.098
GB	0.621	0.651	0.631	0.663	0.651	0.651	0.660	**0.678**
AB	0.591	0.612	0.594	0.610	0.612	0.613	0.614	0.615
XGB	0.630	0.655	0.625	0.659	0.655	0.660	0.666	0.669
BR	0.630	0.655	0.628	0.658	0.655	0.653	0.653	0.660
ET	0.643	0.655	0.636	0.656	0.655	0.650	0.644	0.641
LGBM	0.620	0.640	0.615	0.661	0.640	0.635	0.659	0.661
CB	0.643	0.664	0.634	0.655	0.664	0.643	0.652	0.665
EN	0.647	0.660	0.648	0.001	0.660	0.649	0.659	0.576
SR	0.623	0.653	0.627	0.665	0.653	0.652	0.662	0.672

most models demonstrated declining trends, reinforcing concerns that excessive duplication may introduce noise. While ET and RF performed well at higher thresholds, 50 and 60, RT remained the weakest across the board.

The table also includes comparative results from different modelling strategies. These span five key setups: using only psychological variables (our baseline), stratified split with singleton classes in test data, optimal oversampling results, the inclusion of gender, and the integration of all personal variables. LR was the top performer when only psychological variables or gender were included. Ridge achieved the highest result in the stratified split, while Lasso led the oversampling condition. MLP and GB each led one configuration, with GB showing its best performance when all personal variables were added.

On the contrary, RT consistently ranked among the lowest performers, and MLP struggled with the full set of personal variables. The overall mean R^2 across all models and conditions was approximately 0.627, with higher variability observed in psychological and personal variable scenarios.

These findings emphasize the importance of careful variable selection and modelling choices. While simpler models may offer easier interpretation, they risk losing predictive capability. Likewise, excessive variable inclusion may reduce generalizability. Thus, a balanced approach is crucial to ensuring robustness and interpretability. By minimizing the number of variables, we can reduce the risk of overfitting and enhance generalizability, but this trade-off must be carefully considered. This is why we present the next Table 3.

A quick look at the table, give us a surprising outcome. First, all the models exhibited an improvement in performance, which demonstrates that the genetic algorithm for feature selection was effective. This improvement is particularly intriguing given that the number of variables included in the models is considerably low. It indicates that even with a simplified approach, the models were able to capture essential patterns and relationships within the data. This finding underscores the potential benefits of simplicity in model design, as it not only enhances clarity but also contributes to more robust performance. The table reveals that, considering all the models, the mean R^2 value was 0.65 with a total of 20 variables included in the model. This indicated a moderate level of explanatory power in the models and suggests that they accounted for a reasonable portion of the variance in the data. However, the introduction of a GA demonstrated a notable improvement in performance. The GA achieved an average R^2 value of 0.68, while utilizing only about 12 variables on average. This improvement highlighted the efficiency of the GA in selecting the most relevant variables, thereby enhancing the model's predictive capability without the need for a larger set of variables.

Note that the best possible outcome for the R^2 value was around 0.73. However, the highest R^2 metric obtained during the previous experiments was 0.69. This result came about when the approach involved including singleton variables in both the training and testing datasets, which could be considered a form of «cheating» as it is not supposed to have in both sets the same sample. By doing this, the models likely benefited from information that should have been kept separate, leading to an inflated measure of performance. Despite that,

Table 3 Comparative analysis of R^2. Performance evaluation of non-optimised approach and genetic algorithm-enhanced model. Best values in each R^2 column are bolded, while best values in each row are underlined

Model	Non-optimised		GA	
	R^2	Features	R^2	Features
LR	0.67898	20	<u>0.70250</u>	13
Ridge	**0.68062**	20	<u>0.70010</u>	11
Lasso	0.67919	20	<u>**0.72737**</u>	12
RT	0.52928	20	<u>0.59350</u>	10
RF	0.66565	20	<u>0.67099</u>	15
MLP	0.67898	20	<u>0.71476</u>	10
GB	0.67844	20	<u>0.70097</u>	13
AB	0.61502	20	<u>0.63193</u>	12
XGB	0.66854	20	<u>0.66904</u>	14
BR	0.66045	20	<u>0.67990</u>	10
ET	0.65577	20	<u>0.68872</u>	11
LGBM	0.66147	20	<u>0.67544</u>	13
CB	0.66470	20	<u>0.69354</u>	10
EN	0.66011	20	<u>0.69011</u>	12
SR	0.67226	20	<u>0.68079</u>	15

we achieved better results by not adding singleton classes during training and by optimizing the number of variables.

We would like to point out that the GA does not determine a fixed set of predictors but instead explores multiple solutions as it selects different combinations of variables depending on the model used. This approach allows for flexibility in identifying the most relevant features for predicting our target variable. The selected variables for each model are presented in the following Table 4. It presents the best individuals obtained from the GA optimisation procedure. Each column corresponds to a different model, while each cell indicates the usage of a variable, with a value of 1 signifying that the variable was used and 0 indicating it was not. The last column displays the usage ratio of each variable, and the final row summarises the total number of variables employed across the models. It is important to note that the table is organized by usage ratio and the number of variables, with the highest ratios positioned at the top and models utilizing fewer variables located on the left, creating an ascending order from fewer to more variables.

Besides, it is important to note that the variables in the table are numbered according to their designation in the questionnaires. To maintain coherence across studies, we have kept these original numbering conventions. Consequently, numbers 15 to 20 are not included in the table, as they correspond to the EngI variable.

From this table, we can observe some interesting results. Two variables stand out as the most significant, v1 and v9, which represent Flow and Flourishing, respectively, as they appear in all the models. Furthermore, other highly utilised variables include v21, v13, and v14, corresponding to Emotional Vitality, Regulation, and Total Emotional Intelligence. In opposition, there are variables that are less frequently used, such as v26 and v6, which represent Total Personal and Organizational Quality and Responsibility. It suggests that these latter variables may hold less importance in predicting EngI.

If we turn our attention to the number of variables utilised, it is quite remarkable that the GA effectively optimized the selection. In the worst-case scenario, it decreased the number of variables from 20 to 15, and in the best-case scenario, reduced it to 10, which is half of the original number. In particular, there are four models that successfully employed just 10 variables: RT, MLP, BR, and CB. In contrast, the models that required a greater number of variables were RF and SR. This fact underscores the variability of the approached implemented.

The results of our models, particularly those from the genetic algorithm for feature selection, reveal that Flow, Flourishing and Emotional Vitality are the strongest predictors of EngI, consistently present across the optimised models. Key variables also include aspects of Emotional Intelligence, such as Regulation and overall Emotional Intelligence scores.

To clarify, Flow describes a state of deep absorption and effortless performance, indicating that high EngI involves not only recognising and managing energy but also reaching peak states. Flourishing reflects wellbeing and personal fulfilment, linking EngI with purposeful living beyond cognitive or emotional aspects. Emotional Vitality and Emotion

Table 4 Best individual obtained from the results of the GA. Each column represents a model. Each cell indicates whether a variable was used (1) or not (0). Last column displays the usage ratio of a variable. Last row is the total variables employed

Var	Model															Ratio
	RT	MLP	BR	CB	Ridge	ET	Lasso	AB	EN	LR	GB	LGBM	XGB	RF	SR	
V1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
V9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
V21	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0.933
V13	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0.867
V14	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0.800
V2	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	0.733
V12	1	1	1	1	0	1	1	1	1	0	1	0	0	1	1	0.733
V7	0	1	0	1	1	1	0	1	0	0	1	1	0	1	1	0.600
V3	1	0	1	0	1	1	0	0	0	0	1	1	1	1	0	0.533
V4	0	0	0	1	0	1	0	1	1	1	1	1	0	1	0	0.533
V23	0	0	1	0	1	0	0	0	1	1	0	1	1	1	1	0.533
V25	1	1	1	0	0	0	1	1	0	0	1	1	0	0	1	0.533
V10	0	1	0	0	1	0	1	0	0	1	1	0	1	0	1	0.467
V22	0	0	0	0	1	1	0	1	0	1	0	0	1	1	1	0.467
V24	1	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0.467
V5	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0.400
V8	0	0	0	1	0	1	1	1	0	0	1	0	1	0	0	0.400
V11	0	0	0	1	1	0	0	0	1	1	0	1	0	1	0	0.400
V6	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0.333
V26	1	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0.333
Total	10	10	10	10	11	11	12	12	12	13	13	13	14	15	15	

Regulation form a core part of EngI. It emphasises the emotional dimension's role in managing energy and promoting healthy environments.

These findings offer more than predictive capability since they give understanding of the psychological foundations of EngI.

4.1 Ethical implications

As the integration of AI into research and decision-making processes accelerates, ethical considerations have become increasingly vital. Although our current study did not utilize advanced AI systems in its main analyses, we recognize the responsibility to engage with the ethical landscape surrounding AI applications, particularly in fields related to human behaviour and psychological assessment.

We are committed to aligning with emerging standards in responsible AI development and use. This includes the ethical auditing of AI systems to ensure adherence to fundamental principles such as fairness, transparency, and the minimization of algorithmic bias [55]. Such auditing practices aim to identify potential risks, such as embedded societal biases in datasets, threats to data privacy, or loss of human oversight, and to guide the development of more equitable and transparent systems [56].

These concerns are increasingly reflected in guidelines such as the American Psychological Association's August 2024 policy statement "Artificial Intelligence and the Field of Psychology" [57], which emphasizes the role of psychological science in shaping ethical AI. The APA advocates centring ethics and human rights, ensuring equitable practices, and addressing bias throughout the entire AI lifecycle, from data collection to deployment. Similar principles are echoed in frameworks like the NIST AI Risk Management Framework [56], as well as in the work of numerous global organizations contributing to a growing consensus on the need for auditable and actionable ethical AI standards [58].

Should future iterations of our research adopt more complex AI methodologies, we are prepared to implement formal ethical review processes that involve multidisciplinary collaboration among technical experts, ethicists, and domain specialists [59]. In such cases, tools like IBM AI Fairness 360 or Google's What-If Tool may be employed to audit for bias and ensure fairness in our models [56].

This ethical commitment is also reflected in our attention to demographic factors such as gender. In line with responsible AI practices, our analysis found minimal and statistically insignificant gender influence on model outcomes. As a result, it is an encouraging result that indicates the absence of gender bias and supports the robustness of our findings.

5 Conclusions

Questionnaires are well-established tools in psychology that have been widely used to assess many psychological constructs, including Intelligence. However, relying on self-reported data can introduce subjectivity and bias, because individuals' responses can be influenced by their self-perception, among other factors. In addition to this, administering and analysing questionnaires can be time-consuming and require significant resources, especially for larger populations. This can limit their scalability in certain settings. With the help of AI-based models, the implementation of these questionnaires can be significantly expedited, in this way enhancing the overall process. Besides, AI models help create a more controlled environment that minimises related factors for a more accurate interpretation of certain key variables.

In contrast to traditional assessments methods, the concept of EngI offers a holistic view of intelligence that considers the energetic nature of human beings integrated with free will and Life Purpose. This concept not only enhances the understanding of intelligence but also highlights the potential to integrate innovative assessment methods to explore these energetic dimensions more effectively.

The purpose of this study is to exploit the concept of EngI and propose an AI model to predict an individual's EngI based on commonly measured factors in companies. This study demonstrates the effectiveness of a machine learning approach in predicting this construct. Through the optimisation of several algorithms to extract-related variables, we have shown that our approach can reach a considerable accuracy in estimating this trait. Predicting EngI could help identify individuals with low levels and enable early intervention strategies to enhance their well-being and therefore performance. This may involve personalised coaching, and interventions focused on developing this trait. Moreover, understanding EngI of employees could help organisations create a more supportive and harmonious work environment, to achieve goals doing a more suitable invest of energy and resources not only in employees but also in the teams involved and to promote health in organizations.

In addition to its effects in job environments mentioned earlier, the development of EngI could have a significant impact on adolescents by enhancing four key meta-competencies involved in personal growth and evolution: self-knowledge, self-esteem, self-regulation, and self-development. By understanding their EngI better, adolescents could gain more tools to face the challenges of the digital era we live in today. Another area for applying EngI is family well-being. We predict that fostering EngI in parents, sons, and daughters could strengthen collaboration, communication, and service within families. This would help create stronger family bonds, improve listening skills, encourage meaningful dialogue, and enhance personal relationships. Other potential areas of application include health and sports, where EngI could reduce the material, emotional, and energetic costs associated with achieving results and maintaining balance.

Our findings highlight the potential of data-driven understanding to support psychological research and provide a foundation for further investigations. The analysis carried out in this study revealed that psychological variables are more strongly correlated with EngI than personal factors. So, enhancing emotional and psychological well-being may be more effective than focusing solely on personal circumstances. After several approaches to preprocess the dataset employed, we concluded that stratified splitting, particularly by isolating singleton classes in the testing set, improved model generalisation and performance metrics compared to random splitting. Nonetheless, the oversampling techniques did not significantly enhance model performance and may have introduced noise or led to overfitting. Also, the inclusion of the gender variable had little impact on model outcomes. Furthermore, using all personal variables resulted in a slight decrease in performance compared to models that focused solely on psychological factors. The GA proved effective for feature selection, achieving an average R2 of 0.68 with only 12 variables, compared to 0.65 with all 20 variables. And the best R2 of 0.73 using a reduced model against 0.69 with all variables. This highlights the potential of the proposed model design. Finally, Flow and Flourishing emerged as the most important predictors of EngI, as they were consistently selected by the GA and in Pérez-Moreiras analysis [1]. It underscores their importance in understanding and developing this particular construct.

In future research, we suggest investigating the impact of key variables such as Flow and Flourishing on EngI levels. Plus, we propose a deeper exploration of other advanced feature selection techniques to further improve model performance and identify additional key attributes. Additionally, future work might assess potential bias in model predictions across demographic groups. A deeper analysis with a larger, more balanced sample would help ensure fairness and improve model reliability. Additionally, we propose incorporating explainable AI techniques such as SHAP or LIME, alongside our genetic algorithm approach, to enhance the interpretability of our models. This combined strategy would allow us to identify the most influential variables and to clarify how the models use them to generate predictions. This is particularly valuable when studying novel and potentially multidimensional constructs in psychological domains. We also consider it highly desirable to expand studies on other psychological constructs such as resilience (ongoing), self-regulation and vitality (concepts that, while substantially linked to Flow and Flourishing, have not been directly addressed). Likewise, in the process of consolidating the construct of EngI, it is necessary to approach its study from a neuroscience perspective.

To conclude, we would like to highlight the practical implications of our work. Since 2015, our research team has been implementing practical applications of EngI Development Programs in organizational settings, personal development, and family environments in Spain [1]. The outcomes observed point to increased personal and professional satisfaction, enhanced resilience in facing challenges, and greater self-regulation among participants.

While the use of predictive tools opens new possibilities, we must also remain cautious about overreliance, especially when such tools may risk reinforcing limiting beliefs or self-fulfilling prophecies. EngI offers a perspective grounded in the belief that every individual has a unique purpose and inherent potential. From this understanding, it becomes possible to design environments that foster human flourishing.

We hope that not only coaching psychology, but also education and all areas related to human talent development will benefit from new tools and approaches that embrace this energetic perspective. Looking ahead, we envision a world by 2050 that is more energetically intelligent. One in which each individual is aware of their essential qualities and life purpose, contributing to a deeper harmony with themselves, others, and the planet.

Acknowledgements The authors extend their heartfelt gratitude to the RH Asesores Improving, Talent Development Institute for its invaluable contributions and support throughout the research process. Also, to the I+D+i Improving Project of Energetic Intelligence. Furthermore, we wish to emphasise that this research was undertaken without specific funding, and the authors conducted and concluded the study independently.

Author contributions All authors have contributed equally to this work.

Funding Funding for open access publishing: Universidad de Granada/CBUA. Funding for open access charge: Universidad de Granada /CBUA.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Ethical and informed consent for the use of the data Ethical approval and informed consent were obtained from all participants in this study. This research complies with the principles and evaluation criteria of the Ethics Committee of the Rovira i Virgili University with code: CEIPSA-2022-TD-0031.

References

1. Pérez-Moreiras, E., 2020. *LA INTELIGENCIA Y EL COACHING ENERGÉTICOS: Una aproximación basada en la evidencia desde la Psicología para el Desarrollo de la Inteligencia Energética (Energetic Intelligence), el Fluir (Flow) y el Florecer (Flourishing) | Energetic Intelligence and Coaching: An evidence-Based Approach from Psychology for the Development of Energetic Intelligence, Flow, and Flourishing*. [https://www.tdx.cat/handle/10803/670608#page=1]
2. Díez Fernández-Lomana, J.C., R. García González, and L 2010. Rodríguez García, *Museo de la Evolución Humana*. S.l. : s.n.
3. Cabañas, J.R., *Psicología*. 2001, Almadraba.
4. López González, L., 2013. *Los orígenes del concepto de inteligencia I: un recorrido epistemológico desde el mundo clásico hasta el Siglo de las luces*.
5. James, W., *The energies of men*. 2015: Read Books Ltd.
6. Spearman, C., *"General Intelligence" Objectively Determined and Measured.* 1961.
7. Spearman, C.: The abilities of man their nature and measurement. Nature **120**(3014), 181–183 (1927). https://doi.org/10.1038/120181a0
8. Binet, A., Simon, T.: Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. Psychosom. Med. **11**, 191–244 (1905)
9. Wechsler, D., *The measurement of adult intelligence*. The measurement of adult intelligence. 1939, Baltimore, MD, US: Williams & Wilkins Co. ix, 226-ix, 226.
10. Wechsler, D.: Non-intellective factors in general intelligence. J. Abnorm. Soc. Psychol. **38**(1), 101–103 (1943). https://doi.org/10.1037/h0060613
11. Kambur, E.: Emotional intelligence or artificial intelligence?: Emotional artificial intelligence. Florya Chron. Polit. Econ. **7**(2), 147–168 (2021)
12. Anderson, J.R., *Cognitive psychology and its implications*. 2005: Macmillan.
13. Gardner, H.E., *Frames of mind: The theory of multiple intelligences*. 2011: Basic books.
14. Skrzypińska, K.: Does spiritual intelligence (SI) exist? A theoretical investigation of a tool useful for finding the meaning of life. J. Relig. Health **60**(1), 500–516 (2021)
15. Zohar, D.: Spiritual intelligence: The ultimate intelligence. Bloomsbury publishing, New York (2012)
16. Emmons, R.A.: Religion in the psychology of personality: an introduction. J. Pers. **67**(6), 874–888 (1999)

17. Sternberg, R., 1985. *Beyond IQ: A triarchic theory of human intelligence.* Cambridge University.
18. Sternberg, R.J.: The theory of successful intelligence. Revista Interamericana de Psicología/Interamerican J. Psychol. **39**(2), 189–202 (2005)
19. Sparrow, P.R.: Gestión del conocimiento, aprendizaje organizacional y psicología cognitiva: desentrañando importantes competencias individuales y organizacionales. J. Work Organ. Psychol. **18**(2–3), 131–156 (2002)
20. Yang, S.-y, Sternberg, R.J.: Conceptions of intelligence in ancient Chinese philosophy. J. Theor. Philos. Psychol. **17**(2), 101 (1997)
21. Cianciolo, A.T. and R.J 2004. Sternberg, *Intelligence: A brief history*. Vol. 12. John Wiley & Sons.
22. Ml, C., 2000 *Beyond boredom and anxiety: experiencing flow in work and play*, Jossey-Bass.
23. McDougall, W., 1933. *The energies of men*. The energies of men. Oxford, England: Scribners. 395–395.
24. Feynman, R.P.: The Feynman lectures on physics. (No Title) **1**, 46 (1963)
25. Williams, Z., 2021. *Daniel Goleman's emotionally intelligent contribution to organizational theory.* Journal of management and innovation, **7**(1).
26. Zohar, D., Marshall, I.: SQ: Connecting with our spiritual intelligence. Bloomsbury, New York (2000)
27. Park, P.S., P. Schoenegger, and C. Zhu, *Artificial intelligence in psychology research.* 2023.
28. Salah, M., Al Halbusi, H., Abdelfattah, F.: May the force of text data analysis be with you: unleashing the power of generative AI for social psychology research. Comput. Hum. Behav.: Artif. Hum. (2023). https://doi.org/10.1016/j.chbah.2023.100006
29. Chen, D., Liu, Y., Guo, Y., Zhang, Y.: The revolution of generative artificial intelligence in psychology: the interweaving of behavior, consciousness, and ethics. Acta Psychol. **251**, 104593 (2024)
30. CN, M.A. and S. MR, *Humans Psychological Traits Prediction from their Spending Patterns using Machine Learning Algorithms.* 2024.
31. Baessler, J., Scharzer, R.: Spanish Adaptation of the General Self-Efficacy Scale. Ansiedad Estrés **2**(1), 1–8 (1993)
32. Sanjuán, P., Pérez, A., Bermúdez, J.: Escala de Autoeficacia General: Datos psicométricos de la adaptación para población Española [The General Self-efficacy Scale: Psychometric data from the Spanish adaptation]. Psicothema **12**(supl. 2), 509–513 (2000)
33. Serrano-Fernández, M.J., E. Pérez-Moreiras, M. Boada-Cuerva, J. Assens-Serra, and J. Boada-Grau, *Personality, self-efficacy and self-esteem as predictors of psychological well-being of workers: the flourishing scale (5-FS).* International Journal of Business Environment, 2025. [https://doi.org/10.1504/IJBE.2023.10058504]
34. Biswas-Diener, R., *The new (smiley) face of economics.* 2008.
35. Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., Biswas-Diener, R.: New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc. Indic. Res. **97**, 143–156 (2010)
36. Jackson, S., B. Eklund, and A. Martin 2012, *The Flow Scale Manual.* Mind Garden, Inc.
37. Godoy-Izquierdo, D., M. Vélez, Z. Rodríguez, and M. Jiménez, 2009. *Flow in sport: Concept, evaluation and empirical findings.* XII Andalusian Congress of Psychology of Physical Activity and Sport.
38. Martín-Albo, J., Núñez, J.L., Navarro, J.G., Grijalvo, F.: The Rosenberg Self-Esteem Scale: translation and validation in university students. Span. J. Psychol. **10**(2), 458–467 (2007)
39. Atienza, F., Pons, D., Balaguer, I., García-Merita, M.: Psychometric properties of the satisfaction with life scale in adolescents. Psicothema **12**(2), 314–319 (2000)
40. Vigil-Colet, A., Morales-Vives, F., Camps, E., Tous, J., Lorenzo-Seva, U.: Desarrollo y validacion de las Escalas de Evaluacion Global de la Personalidad (OPERAS). Psicothema **25**(1), 100–107 (2013)
41. Barrios-Choplin, R., Atkinson, M.: Personal and organizational quality assessment revised 4 scale 52 questions. Boulder Creek CA Inst. HeartMath. Retriev. Feb. **16**(1), 2017 (2004)
42. Buchanan, T.M., Reilly, P.M.: The impact of HeartMath resiliency training on health care providers. Dimens. Crit. Care Nurs. **38**(6), 328–336 (2019)
43. Fernández-Berrocal, P., Extremera, N., Ramos, N.: Validity and reliability of the Spanish modified version of the Trait Meta-Mood Scale. Psychol. Rep. **94**(3), 751–755 (2004)
44. Kerlinger, F.N., Lee, H.B.: Behavior Research: Research Methods in Social Sciences. McGraw-Hill, México (2004)
45. James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J.: Linear regression. In: An introduction to statistical learning: With applications in python, pp. 69–134. Springer (2023)
46. Han, H., Dawson, K.J.: Applying elastic-net regression to identify the best models predicting changes in civic purpose during the emerging adulthood. J. Adolesc. **93**, 20–27 (2021)
47. Xiang, L., Cai, M., Ren, C., Ng, E.: Modeling pedestrian emotion in high-density cities using visual exposure and machine learning: tracking real-time physiology and psychology in Hong Kong. Build. Environ. **205**, 108273 (2021)
48. Padmaja, B., Prasa, V.R., Sunitha, K.: A novel random split point procedure using extremely randomized (Extra) trees ensemble method for human activity recognition. EAI Endorsed Trans. Pervasive Health Technol. **6**(22), e5–e5 (2020)
49. Hu, T., L. Yang, Y. Liu, X. Yue, and M. Wu. 2023. *Prediction of the positive psychology of college students using LightGBM algorithm based on SMOTE and Bayesian optimization*. in *2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE.
50. Wang, Y 2022. *Personality type prediction using decision tree, gbdt, and cat boost*. in *2022 International Conference on Big Data, Information and Computer Network (BDICN)*. IEEE.
51. Ogunseye, E.O., Adenusi, C.A., Nwanakwaugwu, A.C., Ajagbe, S.A., Akinola, S.O.: Predictive analysis of mental health conditions using AdaBoost algorithm. ParadigmPlus **3**(2), 11–26 (2022)
52. Vergaray, A.D., Miranda, J.C.H., Cornelio, J.B., Carranza, A.R.L., Sánchez, C.F.P.: Predicting the depression in university students using stacking ensemble techniques over oversampling method. Inform. Med. Unlocked **41**, 101295 (2023)
53. Tahir, M., A. Tubaishat, F. Al-Obeidat, B. Shah, Z. Halim, and M. Waqas 2022. *A novel binary chaotic genetic algorithm for feature selection and its utility in affective computing and healthcare.* Neural Computing and Applications. p. 1–22.
54. Xiang, W.-l, Li, Y.-z, He, R.-c, An, M.-q: Artificial bee colony algorithm with a pure crossover operation for binary optimization. Comput. Ind. Eng. **152**, 107011 (2021)
55. Academy, M., *AI Governance in Practice: Strategies for Ethical Implementation at Scale*. 2025.
56. Lacmanovic, S. and M. Skare, *Artificial intelligence bias auditing–current approaches, challenges and lessons from practice.* Review of Accounting and Finance, 2025(ahead-of-print).
57. Association, A.P., *Statement on Artificial Intelligence*. 2025.
58. Altundağ, S.: Artificial intelligence-based audit software: today's realities and future vision. Denetişim **31**, 180–197 (2024)
59. Bharathan, R., 2025. *Artificial intelligence-comprehensive guide for effective algorithm evaluation and SOX–ITGC and business audit toolkit and examples*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.