Minds and Machines (2024) 34:23
https://doi.org/10.1007/s11023-024-09673-1

A Teleological Approach to Information Systems Design

Mattia Fumagalli[1] · Roberta Ferrario[2] · Giancarlo Guizzardi[3]

Received: 12 August 2023 / Accepted: 22 April 2024

Abstract

In recent years, the design and production of information systems have seen significant growth. However, these *information artefacts* often exhibit characteristics that compromise their reliability. This issue appears to stem from the neglect or underestimation of certain crucial aspects in the application of *Information Systems Design (ISD)*. For example, it is frequently difficult to prove when one of these products does not work properly or works incorrectly (*falsifiability*), their usage is often left to subjective experience and somewhat arbitrary choices (*anecdotes*), and their functions are often obscure for users as well as designers (*explainability*). In this paper, we propose an approach that can be used to support the *analysis* and *re-(design)* of information systems grounded on a well-known theory of information, namely, *teleosemantics*. This approach emphasizes the importance of grounding the design and validation process on dependencies between four core components: the *producer (or designer)*, the *produced (or used) information system*, the *consumer (or user)*, and the *design (or use) purpose*. We analyze the ambiguities and problems of considering these components separately. We then present some possible ways in which they can be combined through the teleological approach. Also, we debate guidelines to prevent ISD from failing to address critical issues. Finally, we discuss perspectives on applications over real existing information technologies and some implications for explainable AI and ISD.

Keywords Information systems design · Teleosemantics · Teleological explanation · Design · Explanation

✉ Mattia Fumagalli
mattia.fumagalli@unibz.it

Roberta Ferrario
roberta.ferrario@cnr.it

Giancarlo Guizzardi
g.guizzardi@utwente.nl

[1] Free University of Bozen-Bolzan, Bolzano, Italy

[2] Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy

[3] Semantics, Cybersecurity & Services (SCS), Twente, The Netherlands

Published online: 18 June 2024

1 Introduction

> *"I don't think we did go blind, I think we are blind, Blind but seeing, Blind people who can see, but do not see"* (Saramago, 1999)

In the novel "Blindness", the Portuguese writer José Saramago tells the story of a strange epidemic. Men and women are losing their sight, going blind, one after the other, seeing everything completely white. Trying to get the contagion under control, the government decides to intern the blind people in groups. But soon the situation degenerates and the whole country becomes infected. As is often the case in Saramago's novels, the narrative anecdote sounds somehow absurd, however, as the story progresses, it gives way to very concrete scenes of life that offer room for reflection on actual issues. One of these that is of particular interest to us concerns the fact that the blindness, of which the Portuguese writer speaks, can be interpreted, paradoxically, not as related to a malfunction of the eye, but as dependent on an inability to observe. Here observing can be understood as "paying attention to", *noticing what is "the other"*, in the sense of *taking into consideration,* "seeing beyond what you see", and in contrast with "looking but not seeing".[1] The infected people described in the novel can see and are blind at the same time because they can no longer notice what is "beyond what they see".

This paradox, in our view, is particularly effective in describing what is happening nowadays in the field of *Information Systems Design (ISD)*, where people find themselves increasingly submerged in a kind of *dark pool*, in which one produces, uses, exchanges information and information systems while "losing sight of" *who* is using and producing them, and, most importantly, *why*, i.e. *for what purpose*. A peculiar effect of this situation is the increasing interest of the community in methodologies and techniques aimed at recovering the understanding of designed systems, along with their functioning and aims, (see, for instance, the case of *Artificial Intelligence Exaplainability (XAI)* (Došilović et al., 2018; Adadi & Berrada, 2018; Guidotti et al., 2018)).

This scenario, then, can be traced back to something similar to the white blindness described by Saramago. Engineers create *information systems*, namely *structures of related components designed to support people in specific tasks by organizing and producing valuable information* (Tilley, 2019). These specialists possess the expertise to design and construct systems that are both efficient and user-friendly and in this sense they perfectly see. Paradoxically, however, in line with Saramago's metaphor, their intensive focus on the engineering process, the intricacies of the artefact, and the tasks to be addressed often lead them to overlook the actual end-users, who are going to use the output of their design. Users are people with diverse cultures, languages, backgrounds, ages, biases, needs, abilities, and disabilities. As a result, designers of information systems end up not seeing the person beyond the

[1] For instance, suppose that while travelling one looks outside the train window, their eyes see everything, but if asked to say what they have seen, they realize that they noticed, i.e. observed just a few things.

user, and then *losing sight of the purpose* concerning their product.[2] This leads to a situation where the assessment of the proper functioning of the system becomes ambiguous, as well as the rationale for its intended use.

It is worth noticing that this kind of "blindness" affects also the individuals playing the role of users, who often do not see beyond the used artefact. For example, in our daily lives, we frequently use numerous applications without knowing their authors' intentions, namely losing sight of the designers' purpose(s). These applications may be the product of a large for-profit corporation or a small group of scientists. Their creation can be driven by factors such as profit generation, facilitating the search for information, or perhaps both, and the diverse purposes influencing the design of each solution often align with the multitude of stakeholders involved. Notably, users are often puzzled in articulating the reasons behind an information system's behaviour, thus explaining *why* an information system operates in a certain way (Shin et al., 2022). Google's search engine, for instance, incorporates ratings and images, possibly for an enhanced navigation experience, but probably also to keep users longer on certain web pages. Social networks interaction features facilitate communication, but may also trigger psychological addiction (Hou et al., 2019). Accordingly, the motivations behind certain responses from artificial assistants and the intentions of their initial designers pose pressing considerations. For instance, reflecting on the purpose behind the texts used to train large language models (Jiang et al., 2020) raises the issue of when and why we should place *trust* in the models we use.

In conclusion, the point we want to emphasize is that following the metaphor suggested by Saramago, blindness consists in disregarding the intentions of the agents involved in the process of design and use of information artefacts. This may sound like an exaggeration, but it seems that we are witnessing something similar to the above-mentioned "blindness" when discussing AI *black box models* (Castelvecchi, 2016), whose behaviour is often opaque to the designers themselves.

In this paper, we propose a teleological approach to information system design that emphasizes the importance of observing, namely, taking care of the subjects and the purposes involved in the information system production process, which are deeply entangled with the system itself. Our solution is strongly inspired by a well-known paradigm in the context of information theories, namely teleosemantics (Macdonald & Papineau, 2006). This theory has advantages that it can be applied to any theory about representation and information production. The key role of teleosemantics is to stress and formalize the fact that each information output has core functions that must be validated by considering the explicitly declared purpose of the creator(s) who produce(s) it and of the users who will later deploy it for reaching their own goals. This validation must be performed also considering that any information production process aims at triggering consumers' actions to successfully

[2] Note that in Saramago's novel, people have no name and, during the pandemic, they are just described with general roles or characterizations, such as "the old man" or "the wife of the doctor". Similarly, designers deal with "ideal users", as they cannot be directly acquainted with each individual using their artefact.

address the designers' purpose. The challenge we are facing is that of translating some of the key notions and principles of teleosemantics in the context of information systems design science (Hevner et al., 2004) and use them to protect designers from being affected by a certain – adopting Saramago's expression – "blindness", and, possibly, decrease future risks.

The remainder of this paper is structured as follows. section 2 describes the problem that drove the settings of our approach. In section 3 we present the three research programs, namely, *Information Systems Design*, *Semiotic Engineering*, and *Teleosemantics*, which provide the basis on which we build our approach, introduced in section 4. In section 5 we discuss how we import the main notions and principles of teleosemantics in the context of ISD. Then, in section 6, we introduce some application perspectives by considering some well-known examples of information systems. Finally, in section 7 we analyse some interesting consequences of the adoption of a teleosemantic perspective on trust and related notions, in section 8 we compare our approach with some related works, and in section 9 we reflect on our results, discuss some future research directions, and provide some final considerations.

2 The Non-Fixed Correlation Problem

> *"We run carelessly to the precipice, after we have put something before us to prevent us seeing it"* (Pascal, 1888)

Through his studies on the language and communication of bees, Von Frisch (2014) showed that a bee, finding a source of nectar during a flight, comes back to the hive and points to the other foraging bees where the source is located. This happens via two types of *dances*. The first, aimed at pointing the source of nectar at a near location, consists in drawing horizontal circles "⊂⊃"; the second, aimed at pointing the source of nectar that is located at a major distance, consists in drawing, through a constant movement of the abdomen (*waggle dance*), a shape similar to the infinity symbol "∞". These two dances, given a nectar location, describe, through an inversely proportional ratio, the distance to the location and the direction to reach it. More specifically, the distance is derived from the number of figures the bee makes in a given period, and the direction is derived from a ratio between the axis of the shapes and the position of the sun. In these figures (⊂⊃, ∞), created by a bee for another bee, pointing to something (the nectar location), we can recognise a form of information exchange, which is very similar to what happens between humans, where the agent producing the information, the one receiving it, and the information itself compose what is often considered as *communication*.[3]

However, there is a key aspect that makes the bees' and the humans' ways of exchanging information substantially different. The bees' language is indeed structured according to a *fixed correlation* between the produced information (i.e., the dances) and the *reference* to which the information points (i.e., the location). For

[3] The bees example discussion and subsequent analysis takes inspiration from Petrosino (1999).

this reason, we could also see the bees' interaction as an authentic *transmission* of information. Each bee pointing to a position, indeed, sends information (i.e., a signal) to another bee immune to misinterpretation, because that information is *unambiguous* and *unique* in its meaning. The bee who receives the signal simply takes that information as it is. When information is exchanged between humans, the situation is different. The same signal, in the form of, for instance, a gesture, a word, a statement, or also an artefact (like a tool or an artwork) could mean different things and, at the same time, different signals could mean the same. In human information exchanges, be they pretty or ugly, there is no fixed correlation between the produced information and its reference. From this point of view, to be sure that the right information is transferred, humans must care about the *understandability* of the 'signals' they produce and they must consider the receiver (consumer) as a *constituent part* of the information production process. The consumer can indeed misinterpret what the producer says due to multiple factors, and the latter has to produce and tune their information considering this possibility. This involves an extra effort in the information exchange process, which often becomes a challenge, independently of the context where it occurs (e.g., natural sciences, social sciences, or ordinary conversations).

This structural ambiguity could not fail to have an impact on *information systems design science*, namely the paradigm "that has the goal to extend the boundaries of human capabilities by creating new information technologies" (Hevner et al., 2004). Naturally, from the realization of omniscient and universal oracles to the atavistic desire to predict the future through a master algorithm, the community of designers in this field has to deal with such an inconvenient non-fixed correlation issue. Unfortunately, it seems that so far this issue has never been fully addressed, perhaps because of its high level of complexity. Moreover, in most cases, it seems that a workaround is preferred, mainly consisting of a focus on the product of the design process, namely the information product itself, putting aside its creator(s) and users. As a very intuitive example, take *Facebook* (now *Meta*), in which the legal channels created by the social network may be exploited to propagate fake news for propaganda, so that algorithms that personalize content are trained with fake data, enlarging and amplifying, even unintentionally, falsehoods, myths, distortions, and half-truths (Mustafaraj & Metaxas, 2017). As a consequence, the social machine seems to be not only uncontrolled but also difficult to control and the ambitious edifices of social media end up being built with bricks bearing our signatures: personal data, which are often incorrect and used for purposes that if spelled out explicitly and in an understandable manner, might not obtain the users' consent for their use.

In a sense, it is as if the output of the design process would extend beyond the humans who create and use it. This is connected to the aforementioned non-fixed correlation challenge. Often, the human producer(s)/designer(s) neglect the consumer(s)/user(s) role as key constituents of the information production process. Acting as a bee, these producers consider the system output information as if it would by itself have a determinate non-ambiguous meaning. As we have already pointed out, unlike the case of bees, this assumption fails in human communication, where the producer(s)/designer(s) owe an *explanation* to the receiver(s)/user(s).

We propose that such a perspective should be especially adopted in the scenario of *Information Systems Design (ISD)* but also in the domain of *Artificial Intelligence Exaplainability (XAI)* (Guidotti et al., 2018), where a multitude of conferences, papers, and research projects are focused on recovering the understanding of information engineering processes and outputs that are nowadays increasingly considered as parts of *black box models*. But, as we will show, this is not the direction taken so far by research.

3 Research Baseline

Before delving into the description of our approach, let us introduce the three research areas from which we take and rework our concepts.

3.1 Information Systems Design

The research about *Information Systems Design (ISD)* has a long history and is rooted in the sciences of information engineering (Simon, 2019), a problem-solving paradigm aimed at guiding the design, testing, and use of information systems (ISs). The IS community widely recognises ISD (Wasson, 2005; Tilley, 2019; Baskerville et al., 2018) as having a key role in the creation and implementation of ISs. In this setting, an *information system* is defined as an artefact, i.e., an *information object*[4] "involving humans and computers, designed to support humans in specific tasks". Such an *artefact* has the *key goal of transforming data into outputs that are valuable to users* and is *always produced through a design process* by starting from the steps composing the classical information systems development life cycle (McDowall, 1991). Examples of information systems applications include *search engines*, *order processing*, *social networks*, *inventory control*, *human resources apps*, *accounting apps*, *e-commerce apps*, and more. These technologies are essentially tools employed in daily tasks, enabling activities, such as making informed decisions, identifying trends, accessing information, and uncovering patterns.

According to the ISD literature, two production levels of information systems can be usually recognised. The first concerns the representation of the system itself, namely the process aimed at describing and identifying all the features, functions,

[4] Though we are aware that these terms are often used in a heterogeneous way within the IS community, from now on we will use the notions of "information system", "information artefact" and "information object" interchangeably, to refer to how information is structured and organized for use in a given application. For a detailed analysis of these notions, we refer the readers to Floridi (2014, 2019, 2023). Also, we would emphasize another interpretation, which may be useful in better understanding our approach, namely that introduced by Ron Weber (1998), who considers an IS as "an artefact that represents entities in a given subject domain, including the states and changes of these entities, and offers IS operations to reflect these changes". The information structures inside these systems reflect a subject domain that is needed to achieve certain goals ascribed to (typically idealized) users of the system. Notice that this definition connects to the fact that in ISD the non-fixed correlation issue relates somewhat to how human communication "implements language in software".

and components of the system. This level involves conceptual modelling activities, relying on design languages (e.g., UML – *Unified modelling Language* (Booch et al., 1996), or BPMN – *Business Process Model and Notation* (White, 2004)).[5] The second concerns the implementation of the system, namely the execution of the design output, i.e., the system representation, through its realization in terms of software and hardware. The output of the former level concerns the backbone structure of the whole system. In this sense, the process of producing an information artefact representation is preparatory to the implementation process, and the output model is what is used to develop, test, and analyse the information system itself.

3.2 Semiotic Engineering

Our proposal is also naturally inspired by *Semiotic Engineering* (De Souza, 2005). This approach was started in the context of *Human-Computer Interaction (HCI)* (Preece et al., 1994) research, but took a quite new path. The main and novel concern of semiotic engineering is that of considering information systems as the means of a communication process between designers and users, rather than only between information systems and users. In this sense, we consider semiotic engineering a natural baseline work from which our solution stems. The primary role of this theory is to bring together the three sources of interpretation and communication involved in the design of interactive computing artefacts: *designers*, *users*, and *computer systems*. Accordingly, the production of each artefact involves, as the first mandatory step, the understanding of what a designer wants to communicate to a user via the produced artefact, and then, as a second step, the design of the artefact itself. Such two-steps process allows for better human-computer interaction and alignment between the designer's and the user's goals.

By considering the communicative context between designers, users, and computer systems, semiotic engineering tries to bridge gaps between other fragmented HCI theories and approaches. Essential to our proposal is the recognition that in the development of the artefact, both designers and users play pivotal roles as active interlocutors during human-computer interaction. This acknowledgement is rooted in the understanding that, through the system's interface, designers provide – directly and indirectly – guidance to users on how to interact with the system to accomplish a specific set of goals envisioned during the design phase. In the semiotic engineering setting, this can be achieved by creating specifically designed spaces for communicative exchanges that users can entertain with the system. These spaces include various elements such as icons, buttons, menus, and other visual cues and features that allow users to interact with the system in a meaningful way. For example, consider a website that allows users to purchase products online. The designers of this website would create an interface that includes elements such as a search bar, product categories, a shopping cart, and a checkout process. These elements are created to be intuitive and easy to use so that users can quickly find what they are looking

[5] Notice that this phase involves also a *requirements engineering* phase, which aims at designing requirements for the system to be in a way that reflects the goals of stakeholders.

for and complete their purchase without frustration. In addition to this, throughout the design process, designers may also conduct user testing and account for features that are required to gather feedback from users, helping the artefact to meet those user's needs and preferences.[6] This involves the opportunity to evolve the artefact itself, embedding new functions discovered by the users, which were not previously considered by the designer.

From a practical perspective, in the specific context of semiotic engineering, emphasizing the importance of interaction and user needs and preferences during the design process helps to create information artefacts that play the role of information carriers or means of communication between designers and users.

3.3 The Teleosemantic Model

The approach we are proposing is also grounded on a well-known information theory, namely *Teleosemantics* (Macdonald & Papineau, 2006). Such a theory provides the concepts that are germane to our proposal and can be leveraged to guide and assess ISD by considering the aim, structural ambiguity, and fallibility of any design output.

Teleosemantics provides an account of information and its production by focusing on the notion of *function*[7] and those of *producer* (or sender) and *consumer* (or receiver) devices.[8] According to this theory, any system is composed of a producer and consumer device, which are equipped, in turn, with specific functions. The information then has to be taken as a state shared between those devices or, in other words, a state to which both devices participate (Macdonald & Papineau, 2006; Giunchiglia & Fumagalli, 2016; Fumagalli et al., 2019, 2020). In the generation of information, the function of the producer is always to generate a state (the *information* itself) according to a certain situation and the function of the consumer is to act in a certain way when the information created by the producer has been received.

Differently from more classical approaches, in which the content of information is seen as its objective referent, according to teleosemantics, the information content is determined by a *success condition*, i.e., the condition that explains if a consumer acts successfully once the produced information has been received. This situation can be easily illustrated by the aforementioned example of the bees depicted

[6] Feedback is not necessarily limited to a specific phase of user testing; ideally, artefacts should be designed in such a way as to foresee a continuous communication process between users and designers through the artefacts' whole life-cycle.

[7] We do not intend here to delve into the extensive debate regarding the notion of function adopted in teleosemantics and the analysis of related key concepts, such as *dispositions* and *affordances*. What we do want to point out is that, in the teleosemantics setting, a function of x can be generally taken as a *capability*, i.e., what enables x to do something according to a certain goal or purpose Cummins et al. (2010), Bedau (1992), Neander (1991).

[8] Here the notion of *device* has to be taken in a broad sense, including, for instance, the sensory apparatus of human beings, animals or artificial agents, or, according to a coarser level of granularity, complex organisms, and information systems. For the sake of simplicity, in the following, we may use "producer" and "consumer" to indicate producer and consumer devices.

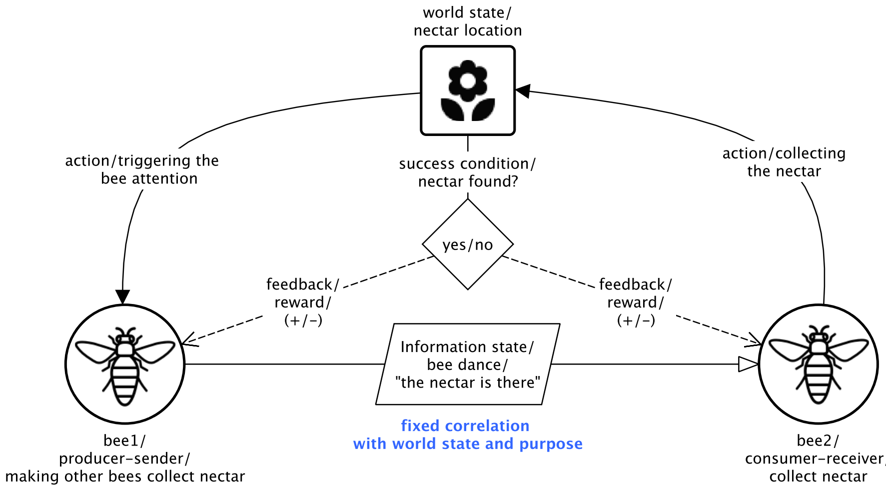

Fig. 1 Overall view of teleosemantics core notions. The bee case

in Fig. 1. Like many other animals, bees produce signals to inform other individuals of the same (or other) species. In Fig. 1 we have a bee (producer) that has the role of helping other bees in reaching a spot with nectar. The producer bee makes a special dance describing the distance and the direction to reach the place. The consumer bee (or some device within the bee) acts for the interpretation of the producer bee's dance. The information is the bee's dance itself. In the example, the function of the producer is to generate a dance in certain circumstances and the function of the consumer is to interpret it and reach the nectar place. Since reaching the nectar place is only successful when the nectar is found (when there is no nectar the trip is only a waste of energy), this is the success condition of the actions/behaviour prompted by the bee's dance. Consequently, the bee's dance *represents* something like "the nectar is exactly there" (Millikan, 2005).

That being said, a key observation is that a "reward" represents what the involved bees obtain when their interaction is successful. This also accounts for the likelihood of this same strategy being adopted again in the future. In these terms, explaining and understanding some produced information amounts to *considering the state of affairs that explains when the actions generated by the information are successful*.

A teleosemantic model can be then described more precisely as follows (Fumagalli & Ferrario, 2019):

An informational state *R* has content *x iff*:

- *R* can be defined as input/output in a system consisting of a producer (device) *P* and a consumer (device) *C*;
- The function of the producer *P* is to produce *R* to obtain *x*;
- *x* is the success condition, dependent on *R*, of *C*'s action caused by *R*.

In summary, teleosemantics can be taken as a way to explain the meaning of representations by tracing back to the purpose of the devices that produce them. This purpose can be more or less successfully realized, always because of the device that consumes/uses the representation. It is important to observe that in the case of organisms such as bees, the representations are mapped one-to-one to what they refer to. This results in an always clear identification of the purpose and, subsequently, the success condition. Differently, in the case of human agents, as we previously discussed, the same representation may refer to more than one thing and vice versa (non-fixed correlation). To trace back to the purpose one must unravel the intentions of the producer. To satisfy the success condition, one must understand the consumer's intentions. In this sense, the meaning of the representation is identified through a process of negotiation between producer(s) and consumer(s).

4 Explaining 'Why'

Much of the effort in IS development has concentrated on enhancing the performance of the tools that implement various functionalities. The ISD community is actively engaged in constructing system models and devising tests to quantitatively assess the performance of such tools. In addition, users and institutions often seek explanations regarding the operation of these tools to gauge their *trustworthiness* (Chen, 2021). As a result, a common practice is to design and adopt simplified models, in particular, to represent and explain complex ISs or IS solutions that carry out tasks without disclosing their internal workings. Primarily, this approach assumes that prospective users are already familiar with such simplified models, which is often not the case. Secondly, these simplified models appear to be constructed *ex-post*, casting doubt on their alignment with the original IS models. Furthermore, while it is essential for all users to grasp the rationale behind design decisions and identify the individuals responsible for creating and maintaining an IS, only a minority possess the technical expertise to understand its intricacies and verify its performance in specific tests.

According to a broader perspective, while information systems can be seen as versatile tools capable of addressing a myriad of tasks, their functionality fundamentally relies on an information exchange process. Naturally, the suitability of the tools used to implement ISs functionalities varies depending on the task at hand and the users involved. Accordingly, considering designers, goals, and users as constitutive and interdependent elements of the information production process, and focusing on their *dependencies*, implies approaching human information products by also taking into account their structural ambiguity and fallibility. This means facing seriously the non-fixed correlation challenge. In this sense, to represent and understand how humans' information design processes work, it turns out to be particularly useful to adopt a kind of *explanation that appeals to purposes*. Through this kind of explanation, something is explained by focusing on its ends, and the functions it should satisfy. For example: eyes have ciliary muscles *because* they allow the eye to focus, chairs have a backrest *because* it allows being comfortable and relaxed while sitting, search engines have a search bar *because* this allows to search for information via natural language queries, emergency vehicles use sirens and flashing lights *because* they signal their presence and make their way through traffic, thus reducing travel time, etc.

Explanations in this form have always been adopted in scientific practice (Keil, 1992; Bedau, 1992; Cummins et al., 2010; Wright, 1976). However, in some contexts, they have been considered as not robust enough. These explanations seem, indeed, to revert the temporal order between the *explanandum*, namely the sentence describing the phenomenon that requires to be explained (e.g., “emergency vehicles use sirens and flashing lights”), and the *explanans*, namely the sentences used to explain that phenomenon (e.g., “this is to signal their presence and make their way through traffic, thus reducing travel time”). The events mentioned by the *explanans* follow the events mentioned in the *explanandum*, while it should be the opposite. This could be quite problematic in the context of natural sciences, where phenomena are usually explained by looking at their causes, which always precede their effects. The central role of the identification of the causes, indeed, is strictly related to the importance of having an explanation that can support the prediction of a given phenomenon. In natural sciences, we can state to understand a phenomenon if we can predict (and control) its manifestation. Nevertheless, in contexts involving purposes that were designed by humans, this is not an issue. The purpose, indeed, can be taken as the reason why a certain human product is characterized by certain features and has certain functions (e.g., everyone would agree that a scissor’s features are a result of the intended purpose of cutting things).[9]

5 Towards Teleosemantic Design

Teleosemantics belongs to the family of teleological explanations, where things and their characteristics are explained through their purpose. In this sense, our approach, which is grounded in teleosemantics, does not aim to formulate a novel theory of information systems design. Instead, it is focused on explaining *why* the products of the design process could be wrong, false, or mistaken and *how* to prevent failures and mistakes (still, by doing so our approach could influence how methodologies for

[9] Note that, according to a standard interpretation of teleosemantics for artefacts, function ascription should ideally occur at the design phase and mainly refers to designers’ intentions Artiga (2023). Still, ISs are significantly more complex than everyday artefacts, such as scissors or chairs. They serve multiple purposes and offer users greater flexibility in how they are deployed and used. Notice also that, in simple cases like that of scissors, a system’s view includes further factors that may go beyond intended functional considerations, especially regarding trustworthiness; these include for instance selling scissors with some protections, not selling scissors to children, etc. In the following sections, we present our approach, based on the teleosemantics framework, in which the set of functions attributed to an artefact represents the outcome of a negotiation process between producers and consumers. Notably, this negotiation can also occur during the artefact’s production, or while it is in use. When negotiation happens in the post-production phase, the addition or removal of functions may result in the design of a new artefact. Whether this newly configured artefact is considered merely an updated version of the original one, or an entirely new creation poses an ontological question that we will not address in this discussion.

ISD are conceived). Additionally, it *makes explicit* and *documents* the main actors involved in the artefact's life cycle and the functions it is expected to satisfy.

From this perspective, our proposal can be employed to analyse the information artefact design process and underscore a crucial aspect: for an information artefact to be assessed as a valid or flawed design output, it must be directly aligned with the *purpose*, or *telos*, of both its designer and users. Otherwise, though it may be a working or somehow useful design output, it cannot be validated and explained. This focus on the designer's purpose, in the context of human information, which is always characterized by structural ambiguity (non-fixed correlation), naturally leads to considering also the consumer(s)/users(s) purposes in the design process. We can thus explain what a wrong or "malfunctioning" information design artefact is, given the *views* of the involved agents. For instance, one can follow their favoured approach to creating an innovative, high-performance artefact, such as a new search engine that aims to provide data to users to organize trips. A teleological approach to ISD proceeds by "explaining" when the artefact is wrong given the perspective of the producer(s) of that artefact, but always in connection with the interpretations of the final consumer(s). Accordingly, the designers and the consumers of the information artefact are kept in the loop of the development life-cycle. Note that this has several implications.

One is that the focus on designers' purposes allows for making explicit "who" provides the requirements for designing the artefacts and "why". For instance, we may have a domain expert, a practitioner, or an individual with a technical background. The primary purpose guiding these requirements might then revolve around objectives such as 'enhancing speed' or 'increasing information accessibility'. A second key implication that needs to be highlighted, which is related to the previous one, concerns the possibility of making a distinction between the designer's and the consumer's purposes, or, also, the designer's requirements and the consumer's requirements. These indeed are not always aligned, rather they can be sometimes conflicting (Davis et al., 2006). For instance, suppose that the main purpose of Google's designers is collecting data and generating profits from advertising, this is not always aligned with the consumer's purpose of finding useful/meaningful information. The distinction between the different purposes of the different stakeholders involved triggers a sort of *meaning negotiation* process, which, seen in the teleosemantics perspective, translates in a *success condition negotiation*, where the designer usually interprets some requirements of the consumers (who may not necessarily be fully aware of what they need) and, by also considering some other design goals, creates the artefact accounting for some specific product requirements. For instance, a consumer/user may require a database schema for querying some information and the designer may go beyond these requirements by proposing an *ontology* on top of it,[10] to enable data

[10] In computer science the notion of "ontology" has a specific technical meaning, which refers to a theoretical or computational artefact used to represent the knowledge managed within a system, to enable reasoning services and to ease interoperability (the exchange of information preserving its content) between systems. For more information about this, we refer the reader to the work presented in Guarino et al. (2009).

interoperability and to allow the user to query the data better. A third implication is that the function(s) of a design artefact, in this perspective, cannot be considered only from the user's perspective. This means that a designer can never be considered merely a passive executor, producing the artefact just following some requirements provided by experts, customers, or users. On the contrary, designers have always their purpose(s)[11] and this deeply affects the production process and the function of the designed artefact. Of course, this does not exclude the common scenario where the designer starts from some inputs from the consumer/user. Still, those inputs need to be considered complementary to the designer's goals, which might not map one-to-one to the user's requirements.

5.1 Information Artefacts as Information Signs

Teleosemantics belongs to the field of information theories and is generally employed to elucidate the involvement of organisms and the environment in the processes of communication, sharing of signs, and representations. Importing the conceptual apparatus of teleosemantics into the context of ISD means emphasizing how information artefacts produced by designers can always be taken as information carriers. *The artefact is always designed according to functionality, but functionality has always a definite meaning*. Moreover, the design process always occurs according to the already mentioned two-part equation that involves both designers and users.

In this sense, we can never assume there is a single, true, direct translation between the functions/meanings intended by the designer and the functions/meanings ascribed by the user.[12] Additionally, in every artefact production process, the designer-consumer dialectic that becomes explicit with user feedback is constitutive of the iterative process of constructing the artefact. Designers' intentions are embedded in the produced artefact but remain open to interpretation and alternative uses (misuses and/or unexpected uses), much like how language can be misinterpreted (or interpreted in new and unexpected ways), leading to unexpected new meanings and translations.

An ISD approach enhanced by teleosemantics takes always the negotiation between designers and consumers as constitutive and allows the artefact to evolve and accommodate artefact repurposing. Consequently, the process of information exchange would not be seen as a straightforward one-to-one transfer from producer to consumer. Instead, it would involve a sequence of interactions that gradually build

[11] Obviously, this is an oversimplification: the designer's purposes may, and often do, include purposes inherited from other stakeholders, like for instance the company in which they work.

[12] To further clarify this point, we want to emphasize that, in the context of ISD, non-fixed correlation might occur for three reasons: (1) the designer does not have direct access to all users. So, their conceptualizations do not perfectly overlap, although they must partially overlap (like in all communications and interoperability tasks communication is a type of interoperability); (2) the designer cannot know all goals of the user; (3) users can project their intention onto the artefact because they can see functions in the artefact that were not designed as such. Sometimes they are not even the users for which the artefact was designed.

and change over time through re-interpretation, ultimately leading to the emergence of new uses, functions, and meanings. This perspective emphasizes the dynamic nature of information artefact design, where the cooperation between designers, users, and the evolving interpretations of artefact use shape the trajectory of the artefact functionalities. By embracing the inherent variability and adaptability of interpretation, our solution aims to offer a framework that encourages innovative and unforeseen directions in the evolution of the artefact itself, while keeping the whole design process transparent and understandable.

5.2 Design Flow

Adopting the perspective of teleosemantics, the design process occurs as a continuous negotiation, in which the designer may want to persuade the consumer to use the artefact according to the design purpose, while the consumer expects to use the artefact to accomplish their own goals. This process is always situated in a context, characterized by some recurring steps and constitutive elements.

Figure 2 provides a visual representation of the process, through a flowchart notation. The design can fail and therefore have a diagnostics and repair strategy associated with it. The process can be iterated multiple times.

(1) *Producing*. The designer gives as output an information artefact, designed according to certain purposes, and associated with a set of functions.
(2) *Consuming*. The consumer takes as input the information artefact and uses some of its functions, to perform different actions, according to certain purposes.
(3) *Assessing*. The information artefact functionality is verified. Both the designer and the consumer can assess whether its application turns out to be successful or not.
(4) *Comeback*. The designer has the opportunity to adapt (e.g., *relaxing* and *constraining* in Fig. 2) the functions of the information artefact, according to the consumer's usage/feedback, to better succeed in a possible next round, and the consumer can decide whether to *continue to use* or not the given artefact.

The above steps have a high level of generality and can be applied to any information production process. Next, we provide more details regarding the elements involved in the process.

5.2.1 Agents and Roles

Any teleosemantic design process requires always a set of agents $a_1 \cdots a_n \in A$ and information artefacts $i_i \in I$. For every information artefact two agents a_p and a_c are considered and one is assigned the role of *producer* (or *designer*), the other of *consumer* (or *user*). Note that those involved in the process are not necessarily single agents, but can

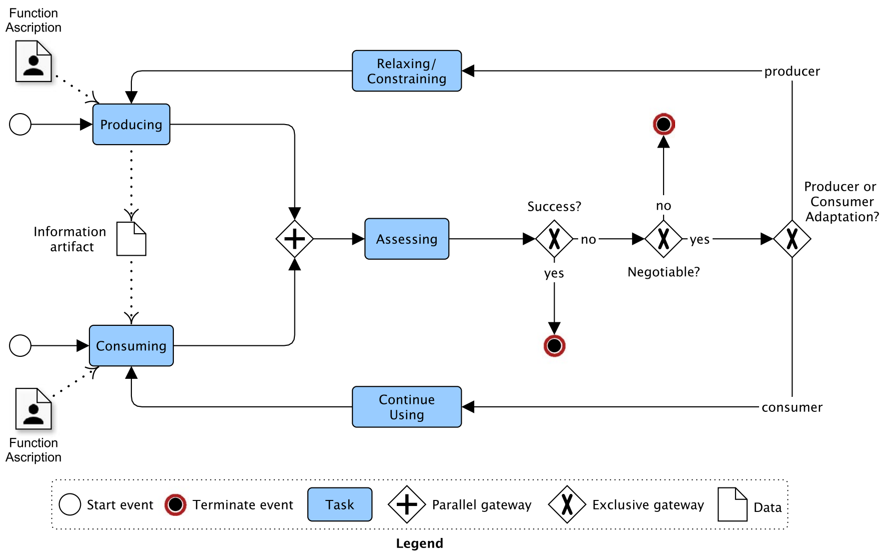

Fig. 2 A flowchart representation of the information artefact production process within the teleosemantic perspective

be also collectives of agents,[13] like, for instance, organizations or business companies, and the role of the consumer can be ascribed to a certain type of user, profiled according to their age or interests.[14] This may already provide useful information about the type of goals the producer may have and the possible needs of the consumer.

5.2.2 Telos and Functions

From a teleosemantic perspective, ISs are always characterized by a set of functions (at least one), which can be traced back to a reference telos (or purpose). The notions of function and telos are in our view deeply entangled and their identification is a prerequisite for verifying the effectiveness of any design output. In our setting, a function can be taken as a particular kind of property associated with the artefact. Such a property, unlike the properties that are inherent to the artefact itself (e.g., the shape of the artefact, some specific components, the structure of the interface, fonts, colours, etc.), depends

[13] In this case, one can refer to techniques of *judgment aggregation* to single out a collective attitude as the outcome of prior negotiations among multiple agents. See, for instance, Porello et al. (2014).

[14] If, on the one hand, the producer is always represented by an existing organization or individual, on the other hand, during the design phase, usually a prototypical consumer is considered. However, upon actual use of the artefact, feedback is obtained by observing the satisfaction or dissatisfaction of the success condition, or, more directly, from individual consumers, who are real persons and organizations. The process of adaptation benefits from this specific feedback, and this can allow a more refined approximation of the prototypical consumer to real-world use cases, or the designer can choose to "personalize" the artefact in some cases.

on the purpose(s) of the agents designing or using the artefact and can be qualified or disqualified accordingly. For instance, an information artefact like a *blog* or a *social network* may have the questionable function of *spreading bad news*, but this depends on the telos of the agent using it (e.g., *misinformation*). Similarly, a *search engine* may have the functions of *providing content*, *collecting data*, or *generating leads*, but these depend on the purposes of the designers. That being said, each telos or purpose can be associated with a set of functions $f_1, ..., f_n$ and indicates why a given artefact is being designed or used.[15] For instance, considering a *search engine*, providing its telos means thinking of it in terms of a specific purpose: *providing information*. Differently, *saying that an information artefact has a function f_i means that f_i enables the artefact to perform a certain action according to a certain purpose/telos.* In other words, what we mean here for a function is the artefact's feature that allows the accomplishment of a telos.

The design process, then, requires each agent to be able to connect a reference information artefact i with a set of functions $f_1 \cdots f_n \in F$. In practice, this means that any agent involved in the process can *ascribe functions* (Kroes, 2010; Griffiths, 1993),[16] namely provide a series of mapping relations as pairs $\langle f, i \rangle$. Accordingly, each *designer* agent a_p is characterized by a mapping operation, namely $O_{ascribes}(a_p, f, i)$, where the ascription means that function f has been expressly designed by a_p to artefact i. The ascription operation, as from Fig. 2, occurs during the *producing* activity. Similarly, the *consumer* agent a_c is characterized by an analogous operation, namely $O_{ascribes}(a_c, f, i)$. In this latter case, the functions that a_c ascribes to i are those that a_c uses; so, differently from a_p, the ascription operation is performed in a successive phase, when the artefact is used (*consuming* in the diagram). All such functions can be then grouped as *producer functions P* and *consumer functions C*, depending on the agent who ascribes them to the artefact, to be able to relate them and evaluate whether the design of the artefact is successful.

5.3 Evaluation Strategy

The involvement and cooperation of all stakeholders in the process are essential to the teleosemantic approach. The practical implication of this aspect can be seen in the stage where a *success condition* for the information artefact has to be established.

5.3.1 Successful and Unsuccessful Functions

Given two agents, a_p (producer) and a_c (consumer), if they ascribe the same function f to artefact i, i.e., a_p has expressly designed f and a_c uses it, then f is a *successful*

[15] The notions of function and telos have been widely debated in the literature. Moreover, here we take the notion of telos and purpose interchangeably. A detailed discussion of these notions is out of the scope of this paper. For the current work, we took inspiration from the analysis in Millikan (1989) and Artiga and Martínez (2016).

[16] Notice that here we assume that function ascription and assignment are the same acts. A more nuanced interpretation of these notions, the distinction between the two related acts, and the impact this may have on the framework we are proposing is beyond the scope of this article and will be the subject of future work.

function, as it is used for the purpose for which it has been designed, the producer and the consumer can be said to be aligned on how to use the artefact. Differently, there may be different cases of *unsuccessful functions*, in which the design output fails.

We distinguish between unsuccessful functions that have been designed but are not used and unsuccessful functions that are present in or allowed by the artefact, though not expressly designed and are used by some consumers.[17] For the sake of simplicity, let's call the former *unused functions* and the latter *undesigned functions*. Both unused and undesigned functions could be either helpful or malicious for the producer and consumer respectively and this also determines a different reaction strategy.

Let us start with unused functions. These are not used mostly because the consumer is unaware of their existence. But such functions can turn out to be useful for the consumer. Take the case of a search engine with an embedded translation service, which is not used by most consumers because it is not very visible on the interface, but that would be used if more easily available. On the other hand, some of these unused functions (but this can happen also with the used ones...) can be useless or even detrimental to the consumer. An example could be one of the many profiling functions present in social networks that personalize the content displayed to the consumer, but that may also be used to collect sensitive data that the consumer would not like to release. Once the consumer is made aware of the existence of such unused functions, they may decide whether to start and use them (in case they deem them useful), or to continue not to use them if they judge them as useless, or they may even decide not to use the whole artefact anymore if they suspect a maleficence on the side of the producer.

Let us consider, then, undesigned functions. The artefact may possess (or allow) functions that the designer never thought about or figured out. Also, the provider may judge these unintended functions as useful, like a tool for drawing diagrams that some consumers decided to use as a logo drawer, or useless/maleficent, such as when for instance a social network is used to spread fake news. Once the producer has become aware of the existence of such undesigned functions, they may decide, in the former case, to make the undesigned function more accessible, so that other users may benefit from it and this is what in Fig. 2 has been called "relaxing". On the other hand, the producer may decide to forbid the use of functions they judge as detrimental, thus performing the opposite action of "constraining". These two actions are very important, as they are what triggers the evolution of the artefact itself.

So far we have considered cases that show when a single function is successful or unsuccessful, but of course, what is much more interesting is to understand whether a whole information artefact satisfies its success condition or not. For this reason, we are going to introduce now some metrics that could be employed to

[17] Note that some functions, although not expressly designed, may still be functional, useful, and thus, in a sense, "successful'' for certain users. However, in this context, we consider a function "successful" only if it is both designed and used. This definition delimits the proper functioning of the artefact as the result of a function negotiation process.

compare similar artefacts or to fix some thresholds to decide when an artefact is well-designed enough to be kept in use.

5.3.2 Assessing the Degree of Success

To evaluate the degree of success of any design output, we propose a set of measures that are relative to the types of functions an artefact may have. The main intuition is that, given the whole set of functions involved in the design process, namely the designer's functions and the consumer's functions (i.e. P and C), the number of successful (both expressly designed and used) functions denotes the degree of success of the design output.

By taking inspiration from well-known information retrieval metrics (Davis & Goadrich, 2006), we can calculate the *transparency* of an information artefact as the ratio of successful artefact functions over the total number of producer functions (so, also including those unused). This can be seen as a measure to quantify how much the consumer is aware of the possible uses of the artefact. In other terms, transparency refers to making *transparent to the user* the designed functions. Differently, *control* quantifies the ratio between successful functions over the total number of consumer functions (thus, including undesigned functions). Control is supposed to measure how much the producer is aware of the possible uses of the information artefact. In other terms, control refers to the designer restricting as much as possible the unintended uses of the artefact. *Transparency* can be formalized as follows:

$$Design_{transparency}(P, C) = \frac{|P \cap C|}{|P|}$$

and *control* is given by:

$$Design_{control}(P, C) = \frac{|P \cap C|}{|C|}$$

Figure 3 provides a visual representation to better understand how the two equations above can be used.

To provide an example, let us take the case where we have a total number of functions (both producer's and consumer's) that is 10, the producer functions are 4, the consumer functions are 6, and the successful functions (namely those that both producer and consumer ascribe to the artefact) are 3. In this scenario, the transparency value of the artefact will be calculated as $\frac{3}{4} = 0.75$, and the control value of the current version of the artefact will be calculated as $\frac{3}{6} = 0.5$. Furthermore, the gathered information can be used to calculate an overall score for the success degree of the current version of the artefact. For instance, the overall degree of success could be calculated as harmonic means of control and transparency, namely $2 \cdot \frac{transparency \cdot control}{transparency + control} = 2 \cdot \frac{0.75 \cdot 0.5}{0.75+0.5} = 0.6$.

In the lower part of the figure, *(i)* provides an example of *good*, i.e., successful, information design output, where most of the functions are successful, namely both

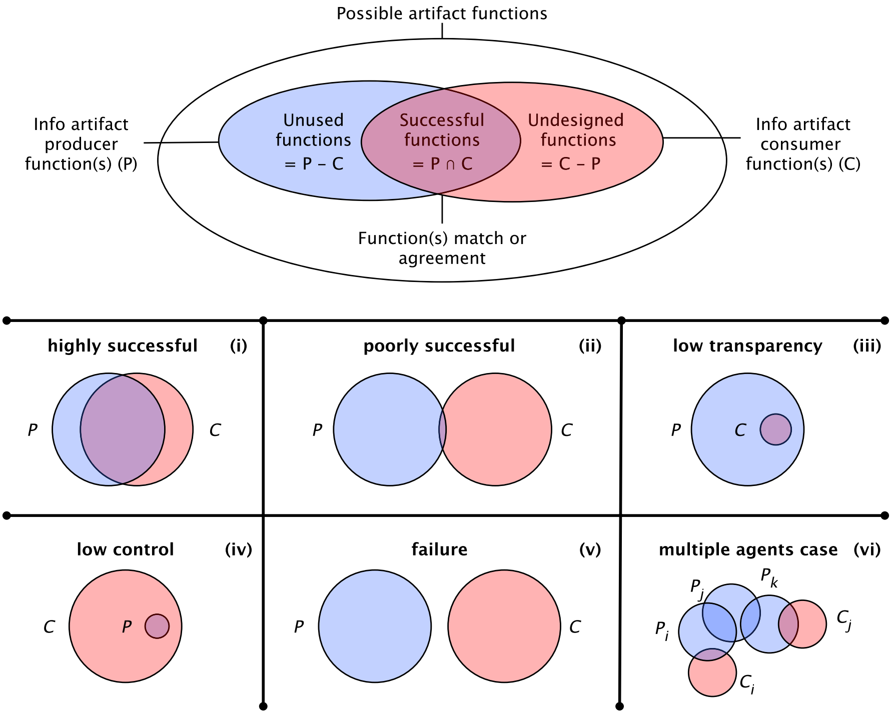

Fig. 3 Different degree of success: some examples

expressly designed and used. In this case, both control and transparency will return high values. Differently, *(ii)* shows a case of bad design output where the producer and consumer share (in the sense of being both aware of) few functions, considering the set of all those available. In this case, both control and transparency will return low values. *(iii)* can be used to depict the case where all functions used by the consumer were foreseen by the producer. However, there is a large set of producer functions of which the consumer is not aware. An example of this situation may be provided by applications that are useful and aligned with consumers' purposes but reuse consumers' data for marketing or profiling purposes that go far beyond the awareness of the consumer. In this case, transparency will return a relatively low value and control will return 1. *(iv)* provides a situation that is opposite to *(iii)*. Here all the functions designed by the producer are used by the consumer. However, there is a large set of functions for which the artefact can be/is used, which were not considered or willingly allowed by the producer. This suggests that the artefact is is somewhat out of of the control of the designer. See for instance cases where social networks are used to spread fake news. In this case, the design transparency will return 1, but the control will return a relatively small value. *(v)* simply represents the case where the design output is a total failure, as no designed function is used by the consumer. *(vi)* represents the case where multiple agents are involved. In this scenario, the values can be extracted by considering the union of the functions from all the agents involved in the process, e.g., $P = P_i \cup P_j \cup P_k, C = C_i \cup C_j$.[18]

Providing additional details about the metrics used for a more elaborated valuation is beyond the scope of this paper. However, as a reference example, it is worth mentioning that for future work we are considering incorporating agents' preferences regarding assigned functions by introducing two refined measures: *average transparency* and *average control*. These metrics, like the ones we have already proposed, are also inspired by metrics in the information retrieval field (Robertson et al., 2010). *Average transparency* can be formalized as $AT@K = \frac{1}{N}\sum_{k=1}^{K} DesignTransparency_k(P_k, C_k) \times successful(k)$. In this setting producers and consumers have to list the functions in order of preference. For instance, the function on top of the list will be the most important. N is the total number of successful functions. K is the total number of functions listed by the producer. k is the rank of the provided function. $successful(k)$ is equal to 1 if the function at position k is successful and 0 otherwise. $DesignTransparency_k(P_k, C_k)$ is the precision calculated at each position, namely, the transparency of the set of *k-preferable functions*. This is based on the ranking orders provided by consumers and producers. Thus, for instance, if we have 6 producer functions, of which only three are successful, the $Design_{transparency}$ would be 50%. However, the $AT@K$ value will vary based on the ranking order. Suppose the successful functions are in positions 1, 4, and 5; in this case, the $AT@K$ will be 70%, namely $AT@6 = \frac{1}{3}(1/1 + 2/4 + 3/5) = 0.7$. Note that the same rationale works for *average control*, which can be formalized as $AT@K = \frac{1}{N}\sum_{k=1}^{K} DesignControl_k(P_k, C_k) \times successful(k)$, where N is the total number of functions listed by the consumer.

It is important to notice that the way the information about the functions is gathered – by observing the use of the artefact, and by collecting *users' feedback* – plays a pivotal role. To explicitly delineate the permissible functions of a given information artefact, while concurrently preventing malicious uses, is not an easy task. Another intricate aspect is to determine all potential uses that were not initially considered in the design, but which should be enabled to better align with the intended purpose. The teleosemantic approach highlights a key feature in this respect, namely, the pivotal role of stakeholders in the design process and their interactive involvement. This virtuous collaboration between producers and consumers is what significantly influences the overall reliability of the assessment phase.

[18] With such union operations, the successful functions would be contained in the intersections between all providers' and all consumers' sets of functions. But it could be interesting to single out the functions that are successful for each consumer, by focusing on the intersection between their functions and the functions of all producers.

It is worth noticing that, to enhance such reliability, functions should be represented in a standardized way.[19] Agents may ascribe several functions and express them by thinking about very general purposes, e.g., *"supporting customer care"*, or more specific purposes, such as *"graph clustering"* or *"community detection"*. Moreover, the adoption of natural language to describe functions may lead to different descriptions of the same purpose. This unavoidable aspect limits the ability to understand whether producers and consumers are pointing to the same, different, or similar functions. To avoid this structural ambiguity, an option is to perform preparatory documentation of functions from which agents can draw. The output of this documentation can be a file with a list of functions that can be gathered from previous well-established analyses, such as that proposed in Blagec et al. (2022) (where a classification of key AI and ISs processes is provided), or that introduced in De Angelis et al. (2023) (where risky functions in the healthcare domain are discussed), or, again, Novelli et al. (2023) (where a general discussion about risky purposes is reported). Given this documentation, both the producer and the consumer can select their intended purposes, still keeping the opportunity to add new ones. The output of the function ascription task is then a list of functions per agent (producer and consumer). There are three main observations here. Firstly, producers and consumers are not necessarily single agents. Both the producer and consumer roles may be played by an organization or a set of agents. The resulting set of functions can be taken as the sum of the functions of the agents involved. Secondly, the process can be applied asynchronously by a producer, given a set of gathered consumer functions, or, *vice versa*, by a consumer, given a set of gathered producer functions. Thirdly, a delicate constraint might be that the involved agents should declare all types of functions, even those that are potentially questionable (for instance, data collection from users should be explicitly declared). The omission of design or use functions can be seen as an act of non-responsibility,[20] on the part of the agents. Finally, the list of functions ascribed by agents has no limit in number. Rather, ideally, the more features ascribed by consumers and producers, the more reliable the validation output. What remains essential is the modelling of a shared list, to ensure proper identification of similarities and dissimilarities among the interpretations of the agents involved.[21] (e.g., *cosine similarity* (Rahutomo et al., 2012)).

[19] Note that the tasks of function documentation and identification, the representation of relations between functions and goals, and relations between functions, have been extensively addressed in the literature, particularly in the context of *requirement engineering* (Yu & Mylopoulos (1998); Eric et al. (2011); Horkoff et al. (2019)), with numerous solutions available. e wish to emphasize that the practical examples provided in the sequel are for illustrative purposes only. The development of a methodology implementing our approach is deferred to future work and will incorporate and reuse relevant existing results from the literature.

[20] The problem of responsibility in designing trustworthy AI artefacts and, more in general, information systems, on the face of the "dark side of technology and its use" is becoming increasingly pressing. In 2022, a special issue has been published on the *European Journal of Information Systems* whose editorial presents a very thorough analysis of the issue Mikalef et al. (2022).

[21] Note that the comparison between the functions listed by the agents can be addressed in several ways. Of course, it can be performed manually, but those lists could be potentially huge and an ideal solution should provide a good balance between the accuracy and efficiency of the comparison process. For future work, we are exploring a *Natural Language Processing (NLP)* approach that transforms each record rep-

With that said, our focus here is on highlighting that this method of quantifying the success level of a design output serves as valuable support for assessing the artefact itself and lays the groundwork for activities related to artefact evolution. Moreover, the ideal scenario in which designed and used functions perfectly align seems to be precluded by the non-fixed correlation problem. However, it is crucial not to underestimate the role played by negotiation and the modifications made by the designer in response to user feedback in the artefact's life cycle evolution. In a sense, the "partial" success of the design output and the subsequent negotiation for desired changes serve as a wellspring of innovation.

6 ISs Under the Teleosemantic Lens

As anticipated in Sect. 5, the teleosemantic approach to information systems design should be taken as a kind of piggyback theory that does not aim to substitute currently existing design theories, but is rather applied on top of them. The key aspect added by teleosemantic design concerns then the design output verifiability. In the teleosemantic setting, indeed, the information artefact, with its structural ambiguity, should always be created and assessed by referring to the designer's purpose and the consumer's use involved in the process. Accordingly, the new issues raised by the proposed approach can be addressed at the design phase, when a well-founded conceptual representation (Guizzardi, 2005; Guizzardi et al., 2010; Guizzardi & Reis, 2015; Guarino et al., 2020; Biccheri et al., 2020) of the information system is produced, but possibly also at run time, to properly gather feedback from consumers.

To discuss how our approach can be used to support information systems design, we selected two examples of information system technologies from different fields of artificial intelligence, one from knowledge representation and one from machine learning. Each solution presents multiple facets when viewed through the lens of the teleosemantic perspective. Our goal is not to perform an evaluation and a systematic analysis of the quality of the selected solutions, but rather to highlight their characteristics through the presented approach.

Some of the issues we address to discuss the examples concern the main purpose and related functions of the given information artefact. Moreover, we focus on the stakeholders involved and the degree of success of the selected solution. These are some of the critical issues to be addressed for verifying an information artefact according to the teleosemantic perspective. Note that, to take full advantage of the proposed approach, purposes, functions, and stakeholders should be explicitly defined. From our perspective, these are the baseline components to *explain* the implications for practice and research of any given design output. Ideally, the framework should be applied at the design phase, so that it can be used constructively, to

Footnote 21 (continued)
resenting a function in the list into a corresponding vector. These vectors can be then matched through a similarity function.

adapt the artefact to the outcomes of the success condition's negotiation process. Nonetheless, it can also be employed at run time.

In the following cases, for illustrative purposes, we briefly reconstruct the elements indicated by the framework *ex-post*. This will bring out more of the critical potential of the teleosemantic approach and less of the constructive aspect.

Schema.org.[22] Schema.org is a graph-shaped data structure consisting of a taxonomy of concepts, characterized by a huge variety of properties that constitute what, according to the knowledge representation jargon, is called *vocabulary* or *knowledge graph*. Still, since its first release (2011), after being adapted and evolved by multiple subjects, Schema.org has been used in many contexts for addressing a variety of goals.

- *Agents and Roles.* The Schema.org project is primarily the output of a community effort. However, the founders of the project can be traced back to big business players, such as *Google*, *Microsoft*, *Yahoo*, and *Yandex* (these companies can be regarded as the designers of the project and assume they have a shared purpose). From a production perspective, the presence of multiple designers allows for the identification of multiple and varied functions and corresponding purposes. Similarly, from a consumption/use perspective, users of the artefact may play different roles, such as project founders, knowledge engineers, or end users of the search engines who benefit from the whole schema.
- *Telos and functions.* The business role of all founders already suggests some directions. On the one hand, as also declared on the project main page, Schema.org has the main mission *"to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond"*. The teleological approach here requires answering a "why-question" and a clear-cut answer is expected. What can be gleaned from the available information is that the whole project is created to optimize some search engine results and related use (see for instance the possibility of visualizing recipe information in the search snippets). In other words, the goal of this vocabulary seems to be that of offering a markup facility for web pages. On the other hand, as expected, the use of this artefact suggests analogous and/or different functions. By scraping available research work (since 2011) we can find different scenarios of application. The largest adoption is for sure structuring web documents. We can find Schema.org as a reference standard to support *job posting* (Guha, 2011), for *hotel booking* (Kärle et al., 2024; Fensel et al., 2016), or, more generally, *website development* (Tort & Olivé, 2014). We also have works where the function of enhancing search engine results is explicitly declared (Sulé, 2015). More recently, Schema.org was used also to enable (Panasiuk et al., 2018) conversational tourism assistants and to train machine learning models for product matching or entity types matching (Peeters et al., 2020; Giunchiglia & Fumagalli, 2020). Naturally, all these functions and related uses can be mapped to different purposes of different agents.

[22] https://schema.org/.

- *Unsuccessful functions.* The assessment of the functions embedded in Schema.org depends on the associated purposes. The selected purpose, then, affects the reliability of any following verification process. For instance, suppose that the purpose of Schema.org is to support Google's users in navigating the search engine and decreasing the *bounce rate* (i.e., the time of stay on Google pages). In this case, all the information encoded by the Schema.org graph should be verified accordingly. For instance, given that purpose, the function of enabling the training of accurate ML models or reasoning about tourism data can be identified as an example of an unsuccessful function, as it negatively affects the achievement of the selected purpose (decreasing the bounce rate).[23] To clarify these issues, an effort to make explicit the purpose of the artefact is required. This also would help developers and designers in better focusing their evolution activities. Once negative feedback from the consumers is received, the designer may choose to *constrain* the functions of its information artefact.
- *Degree of success.* The lack of a clear-cut interpretation of the artefact's purpose and the uses that diverge from the stated design functions naturally suggest that Schema.org is an information output with a *low level of control*. For instance, what if Schema.org vocabulary is taken as an ontology to enable interoperability between applications? The degree of success in this direction would be certainly very low, as the functions to obtain such a purpose were not designed. However, we have a *description logic (DL)*[24] encoding of the resource (Hernich et al., 2015). It seems that we are facing here a case with undesigned but used functions.

ChatGPT.[25] ChatGPT is an artificial intelligence chatbot built on top of the GPT families of language models (Jiang et al., 2020), developed by *OpenAI*. This AI algorithm is based on unsupervised machine learning and works by leveraging a deep learning technique known as "transformer", which involves using a neural network to analyse and exploit textual information. This new technology was implemented and made public in November 2022. Due to its power, ease of use, and the possibility of accessing it for free, its use increased exponentially after it was made public.

- *Agents and Roles.* A key aspect that emerges by analyzing ChatGPT with the lens of the teleological approach is the multitude of roles that can be potentially involved. Firstly, behind the ChatGPT project, we can find the effort of the OpenAI artificial intelligence research laboratory. This organization consists of both for-profit and non-profit corporations and involves many stakeholders. Secondly, ChatGPT is trained over a huge amount of text corpora. Those data affect and

[23] Note that in this sense, undesigned functions are unsuccessful when they conflict with other stated goals (in this case of the producer) and not merely because they were undesigned.

[24] DLs are logics used to represent computational ontologies. For more information see Baader (2003).

[25] https://openai.com/blog/chatgpt/.

bias the way the application works and were created by authors (i.e., designers) who are independent of the ChatGPT project. See, for instance, the Wikipedia content, which is among those used for training ChatGPT. This was created in a completely different setting, by people who did not have in mind the multiple possible ChatGPT applications. Thirdly, ChatGPT is used by people with different expertise (e.g., marketers, engineers, humanists). Moreover, anyone can naturally build apps on top of ChatGPT, thus contributing to its evolution and design.

– *Telos and functions.* The claim on the OpenAI website is surprisingly broad and pretentious: the main purpose, indeed, is identified in the *"development and promotion of new AI technologies in a way that benefits humanity as a whole"*. With such a vague designer's purpose, the identification of users' purpose becomes much more prominent. However, anyone can potentially be a ChatGPT user, and anyone can use the technology in multiple and different domains with different goals. This scenario, while the technology is assessed regarding a specific application domain, involves *altered control and transparency values*. As for the Schema.org case, the teleological framework stresses the fact that the opacity of purpose(s) and roles in the creation of ChatGPT involves a potentially high number of unsuccessful functions, thus requiring a potentially huge effort in terms of negotiation. For instance, if the main telos is traced back to the purpose of the authors who generated the textual data used for training the model, the characterizing functions should be ascribed accordingly. Suppose we stick to the *Wikipedia* authors' purpose, namely allowing people to access encyclopedic (crowd-sourced) knowledge. ChatGPT-designed functions should then enable answering general encyclopedic questions (and not highly domain-specific questions, as, for instance, those related to the medical domain).
– *Unsuccessful functions.* From a teleological perspective, The ChatGPT technology raises interesting scenarios. An exemplary case is provided by the high number of possible undesigned functions. For instance, the disclaimers provided when opening the application (i.e., "may occasionally generate incorrect information", "may occasionally produce harmful instructions or biased content", and "limited knowledge of world and events after 2021") sound problematic. These suggest that the artefact admits a large number of undesigned functions and questions the possibility of trusting the answers provided by the ChatGPT artificial assistant. Since its public release, the application has been subject to several criticisms. Consider, for instance, the many possible "overfitting" cases, where ChatGPT provides answers even if it does not have enough data. The chat, if domain-specific information is asked, may provide wrong answers. However, these answers may seem reliable to users, thus increasing the risk of misinformation (see, for instance, when we ask about a specific concept like *minimal cognitive grid* (Lieto, 2022), or when we use it to solve simple mathematical problems – one of the typical tasks in which, so far, ChatGPT has failed). Even if ChatGPT is designed to learn from its errors, without having a clear understanding of the text corpora used to train it and the purpose/role/reliability of the text's authors, the amount of undesigned and potentially harmful functions is large. Similarly, suppose that ChatGPT-

designed functions are mainly aimed at collecting data from users. This issue has been widely debated and in several countries has led to temporarily blocking the access. One crucial question revolves around whether a policies document explicitly stating that the data will be used for market research and advertising purposes is sufficient to enhance clients' awareness.
- *Degree of success.* At the current state, from a teleosemantic perspective, the ChatGPT design output has still a lot of room for improvement in terms of *transparency* and *control*. For instance, using ChatGPT to support students in solving simple mathematical problems could be an issue. The overall degree of success in this direction would be certainly very low. Moreover, even if the model could be improved in the future, its application in the math education domain may even be discouraged, since its purpose is not to make students learn math. This scenario can be found in many other domains of interest. Users may find out, at their own expense, that the model is unreliable in certain contexts. Alternatively, the designer, by clearly stating the purpose, could promote the correct use of the technology. Most of the questions that are currently widely debated seem likely to benefit from the evaluation method we have proposed.

7 Implications of the Proposed Approach for Trustworthy AI

The teleosemantic framework we are proposing documents and makes explicit all the stakeholders involved in the production of an information system, what such a system is supposed to do, and the functions that have been designed to accomplish its purpose. In this way, the trust[26] of the users on the system should be enhanced, at least because they can tell when it fails and who is to blame for the failure.

Let's try to delve a bit into the idea of trustworthiness and how a teleosemantic approach may have a role in reshaping what is at stake. In the engineering literature, the term "trustworthiness" is mostly interpreted as a synonym of "dependability".[27] The latter term interestingly recalls the idea of instrumental dependence, thus pointing out how humans may depend on the use of a specific artefact for the accomplishment of their own goals or purposes. But if we are to depend on an artefact to accomplish our goal, we would like the latter to be *reliable*, i.e. we would like to be able to predict its behaviour and we would like such behaviour to imply some positive outcomes towards the realization of our goal. In this sense, by spelling out the success condition, teleosemantics may provide a way to test the reliability of information systems and, more in general, artefacts. However, the concept of "trustworthiness" has become central especially when referred to the domain of Artificial Intelligence, where systems are taken to possess a great degree

[26] Notice that, in the philosophical literature on trust, the latter is often seen as a quaternary relation involving a trustor, a trustee, an action, and a goal, see Chen (2021) for a reference.

[27] The reader may find a more thorough conceptual analysis of the notions of trust, reliance, trustworthiness, dependability, reliability, etc. in Biccheri et al. (2023).

of autonomy. Nevertheless, analogously as in the so-called *Hybrid Human-AI Manifesto* (Akata et al., 2020), our approach sees AI as a prosthesis augmenting human capabilities, rather than a substitute for human agency; this means that being reliable is not enough for the AI artefact to be deemed fully trustworthy, as also the humans (individuals or organizations) who use it as a prosthesis to augment their capabilities should be trustable.

The model we propose, by "forcing" all stakeholders to declare their identity, their purposes, and the functions they ascribe to the artefact, makes the whole system more transparent, and therefore more trustworthy.[28] Take, as an example, the case of the co-creation of AI artefacts. Part of this process of co-creation happens consciously when users provide their feedback to designers, who may ameliorate the product or personalize it according to the specific user's desiderata. Nonetheless, nowadays this process of co-creation is, at least partially, hidden. Most users are not completely aware of how their data are used, for instance, to train the ML algorithms, or which are the organizations behind the development of the AI artefact. This strongly hinders their freedom to choose whether to provide their input for the co-creation of the artefact. In fact, in 2019, the European Commission published some guidelines for trustworthy AI; such guidelines are based on five ethical principles: Beneficence (do good), Non-Maleficence (do not harm), Autonomy of humans[29] (preserve human agency), Justice (be fair), and Explicability (operate transparently). The principles are then supposed to be translated into ethical requirements (Guizzardi et al., 2023). By looking at the five principles, one gets the impression that to check whether the former four are satisfied, one needs to enforce the fifth principle and obtain transparent systems, whose actions and decisions can be seen and understood. In fact, explicability (a term introduced by Floridi and colleagues in (2018), to point to a complex concept, including in its definition both intelligibility and explainability) appears to be a pre-requisite for controlling that the other principles are followed. At this point, it is worth reflecting on what we mean when we say that we would like systems that are intelligible and explainable. The first question that deserves a serious answer is: "intelligible for whom? And explainable to whom?"

So far, most studies in explainable AI have been focused on the "translation" of black box models into more human-understandable models, like for instance bayesian networks. But those who can understand such models are, in the best case scenario, users who, for some reason, have studied them, for sure not 'average' users. On the other hand, the basic elements of teleosemantic models are very intuitive and widely intelligible. If from here we turn to explainability, the main point is that classical XAI approaches are meant to *descriptively* explain *how* an artefact does what it does.

[28] In other terms, not only the AI artefact should be trustworthy, but the whole socio-technical system including producer, consumer and all agents involved. Interestingly, the theory of teleosemantics foresees a mechanism of *zooming in* and *zooming out* which allows us to apply the analysis at different levels of granularity.

[29] Notice that to preserve their level of autonomy, humans need to understand the functions of the AI artefact they are using, otherwise the latter is out of their control.

On the contrary, the approach we are proposing aims at providing the consumer/user with a *teleological* explanation, which shows *what* the artefact does and for which *purpose*. We argue that this kind of explanation is well-suited for artefacts, which are built with a telos and used with a purpose, it is more useful and informative for consumers/users and is helpful also for designers.

A notable discussion somehow related to the aforementioned attitude is presented in Mikalef et al. (2022). Here a holistic understanding of responsible AI is advocated, focusing on how to reduce unintended negative side effects while raising customers' knowledge and awareness. The authors of this work describe responsibility – which indeed is a concept that is closely related to trustworthiness – as a multidimensional asset, including transparency, accountability and robustness. In our paper, we have tried to promote the former dimension by reducing unused functions and the latter two (well represented by our notion of control) by reducing unintended functions. A key aspect discussed in Mikalef et al. (2022) consists of translating responsibility dimensions into a rich list of research questions. In our proposal, we have not addressed many of those questions, for instance, those related to human biases, data privacy and governance, or the social or environmental effects of the employment of AI artefacts. In particular, we have not dealt with the enormous problem of unforeseen risks that can bring about unwanted outcomes that only emerge at the run time.[30] At the same time, the teleological approach we propose, which focuses on ensuring the accessibility of functions to users, implementing iterative design changes based on users' experience, and holding the main stakeholders accountable for selecting functions for AI systems, can naturally be viewed as a method of tackling responsibility issues in ISD.

Interestingly, the "Artificial Intelligence Act"[31] of the European Union already highlights some of the issues we have tried to address with our proposal, starting from the idea that AI systems may be classified based on their level of risk (unacceptable, high, limited and minimal), depending on their context of use, as is the case for instance of emotion recognition systems, which may be used fairly straightforwardly for therapeutic purposes but should be considered as high-risk – and thus strictly regulated – when they are used in law enforcement contexts or at the boarders for regulating migration and, in any case, not without notifying the interested person. More in general, the explicit identification of the producer is also taken into account: persons (users) should be notified when they are interacting with an AI system and not with another human, to prevent impersonation or deception. Our teleosemantics-inspired approach already implicitly satisfies such requirements by *i)* representing explicitly the (human or artificial) producer and *ii)* the purpose for which the AI system is employed. This means that users would be "automatically" notified about who is using their data and for which purposes; furthermore, being also the purpose necessarily explicit, the classification of AI systems according to their level of risk should become much more transparent. For all these reasons, we

[30] The reader may find some discussions on this point in Chiffi et al. (2022).

[31] The last available version of the act is the 21st January 2024 draft, see https://artificialintelligenceact.eu/ai-act-explorer/.

believe that working on the implementation of the proposed approach goes exactly in the direction of the enforcement of the AI Act.

8 Related Work

As we have pointed out, our proposal aims to offer a new method of validating information systems. The key premise is that the design process must always involve an examination and understanding of the actors involved (designers and users), their purposes, and how they interpret the functions of the artefact they produce and/or use. In this sense, the teleosemantic approach to design finds an important ally in what has been referred to as *Social modelling*, or more specifically, *Goal-Oriented, Requirements Engineering (GORE)* (Eric et al., 2011; Yu & Mylopoulos, 1998; Horkoff et al., 2019).[32] This area of research offers indeed a new way of tackling the requirements elicitation and definition challenge. Here an information system is primarily taken as a means to improve the interaction between actors. Accordingly, before focusing on the behavioural properties of software, as in a mechanistic system, a higher level of abstraction is required. This involves identifying all relevant stakeholders who will be affected by the system, their high-level goals, and, possibly, more specific sub-goals. Through the systematic representation of these elements, a more comprehensive and accurate derivation of system functional and non-functional requirements should be elaborated. One observation is that GORE processes effectively capture requirements in terms of stakeholders' goals (intentions). However, the goals of the system designer or the organization behind the system are typically overlooked.

An important aspect of Goal-Oriented Requirements Engineering research is that it leverages and extends highly used and studied modelling languages and notations such as UML (Booch et al., 1997), as well as older ones like Data Flow Diagrams (DeMarco, 2001) and SADT (Dickover et al., 1977). Moreover, it has paved the way for the production of a plethora of new standards for goal-oriented requirements representations. Among the best-known works in this direction, frameworks such as the NFR framework (Chung et al., 2000), KAOS (Van Lamsweerde, 2001), and GBRAM (Anton, 1996) have placed goals at the centre of their models. Finally, the *i** framework (Dalpiaz et al., 2016) provides a language to model information systems as artefacts existing within social environments, where actors interact with each other based on goals to be accomplished, tasks to be executed, and resources to be provided.

Overall, this key-related work represents a concrete opportunity to implement and operationalize what we have proposed here. With its new emphasis on "why" questions, functions interpretation, negotiation, and success conditions, the teleosemantic approach can be formally expressed by a (possibly adapted and/or extended) goal modelling technique, thus naturally supporting the determination of systems

[32] Concerning GORE we refer the reader to this recent talk by John Mylopoulos: https://www.youtube.com/watch?v=7zqa2lPeKdw. The talk connects GORE to a style of teleological explanations.

requirements processes, and then assisting in the design and validation of systems that truly meet the stakeholders needs.

9 Conclusion and Future Perspectives

In this paper, we have presented an approach to the analysis and documentation of information systems based on teleosemantics, a well-known philosophical theory on the transmission of information.

The purpose behind the application of such a framework is twofold: on the one hand, it should help prevent incorrect uses of the system (thus, increasing *control* over it) and, on the other, it should favour user's awareness about its functions (increasing *transparency*).

One of the main tenets of teleosemantics is that the content of information is not determined *semantically* by singling out a sort of objective meaning of the information, but is rather established *pragmatically* by referring to a *success condition* for the actions performed by a consumer and induced by the information communicated by the producer.

In the classical examples taken from biology, like that of the bees discussed in Sect. 2, such a condition is one that, if realized, contributes to evolutionary progress. Obviously, the situation is very different with information systems and, more specifically, with AI artefacts, as they involve an exchange of information between humans, for which the non-fixed correlation problem emerges, as illustrated in Sect. 2. In such a case, the producer and the consumer most of the times will not consider the same conditions as successful. For this reason, we have introduced in Sect. 5 a mechanism to negotiate what producer and consumer can agree to call a 'success condition'. The evidence of the success condition is in this case the continued use of the artefact, whose functions can be added or constrained to proceed to a new negotiation.

The motivations behind our proposal to use teleosemantics are many. First of all, the role of the consumer/user is central in the theory, as they play an active role in the determination of the success condition, in such a way that we could identify here a sort of *co-creation* of the content of information systems, thanks to their feedback through use (Lee, 2020). In more technical terms, such co-creation would be the result of requirements elicitation and artefact specification through a continuous and iterated process of user feedback and artefact adaptation on the designer's side. Such a process is at play both in the design and at the run time.

Second, the teleosemantic model offers the opportunity to explicitly document in a clear way what a system does (telos), how it does it (functions), when it works or fails (success condition) and which are the stakeholders involved in its design (producer) and use (consumer), thus enhancing the trust of users and society at large, as illustrated in Sect. 7 for the specific case of AI artefacts.

Of course, we are well aware that proposing such a framework as a good practice is not enough to induce trust in AI artefacts, this would require much stronger enforcement. For instance, one of the possible solutions would be to create a sort of repository where AI artefacts (or more generally ISs) would be listed with explicit declarations concerning all the elements included in the teleosemantic framework. This would bring with it the necessity of serious consideration about the policies for AI, which should be represented in greater detail w.r.t. what is happening now. In a sense, what we are encouraging with this paper is a way of handling the dangers of AI, in line with what happened in the 90s in the food industry, when it became recommended and then compulsory to label products with important information, like ingredients, additives, calories, etc. Then, there should be an international authority monitoring whether what is declared in the repository is actually the case, etc. The AI-Act already mentions a ‘public EU-wide database for registering standalone high-risk AI applications’; one could think about structuring such a database according to the teleosemantic approach and extend its use to all AI applications, also because sometimes dangers emerge when artefacts are used in ways that were not foreseen by designers. All this is evidently challenging, but it is also a matter that concerns politics and law, as well as research. In this paper, we provide a first step toward a direction that we deem more in line with the European Guidelines for a Trustworthy AI and the AI-Act.[33]

To go back to the research, we conclude by saying that the next steps to be considered are a refinement and an extension of our framework. The refinement concerns the evaluation strategy, which has been provided in this paper, but still needs to be practically tested and different open issues require to be addressed and extended considering practical outcomes from related research, such as the one in GORE. At the current stage, for instance, in the evaluation, we directly count functions, but functions can play a more or less prominent role in the overall functioning of an AI artefact, so further parameters with weights, ranking, or probabilities are likely needed for building mathematically more elaborate models.

A final issue left for future work is the inclusion in the model of collective agents, both in the role of the producer and of the consumer. It is not by summing up the designed and used functions that we can single out the success condition of an artefact in a collective scenario, as, for instance, different users may want to use or not use additional functions. Maybe other negotiation mechanisms, “internal” to the collectives of producers and consumers need to be established, but this is undoubtedly an issue that deserves more investigation in the near future.

Acknowledgements Roberta Ferrario was supported by the BRiO Project, funded by the Italian Ministry of University and Research (MUR) under the PRIN Scheme (Project nr. 2020SSKZ7R), by the project SMARTEST (Project nr. 202223E8Y4X), funded by MUR under the PRIN Scheme and by the project FAIR, (Project nr. PE 00000013), funded by MUR.

Author contributions Mattia Fumagalli conceived the initial idea of the paper, designed its narrative structure and led the development of the whole contribution. Mattia Fumagalli and Roberta Ferrario co-wrote the text of the manuscript, Roberta Ferrario focusing on sections Sec 7 and 9. Giancarlo Guizzardi contributed to evolving the initial idea of the paper, in particular w.r.t. connecting the proposed approach to Information Systems Design and related areas (Semiotic Engineering, and Goal-Oriented

[33] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

Requirements Engineering). All authors contributed to the conception of the approach and manuscript revision, and read and approved the submitted version.

Funding Open access funding provided by Libera Università di Bolzano within the CRUI-CARE Agreement.

Data Availability We do not analyse or generate any datasets, because our work proceeds within a theoretical approach.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., … Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer, 53*(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Anton, A. I. (1996). Goal-based requirements analysis. In: *Proceedings of the second international conference on requirements engineering* (pp. 136–144). IEEE.

Artiga, M. (2023). A dual-aspect theory of artifact function. *Erkenntnis, 88*(4), 1533–1554.

Artiga, M., & Martínez, M. (2016). The organizational account of function is an etiological account of function. *Acta Biotheoretica, 64*, 105–117.

Baader, F. (2003). *The description logic handbook: Theory, implementation and applications*. Cambridge University Press.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems, 19*(5), 3.

Bedau, M. (1992). Where's the good in teleology? *Philosophy and Phenomenological Research, 52*(4), 781–806. JSTOR.

Biccheri, L., Borgo, S., & Ferrario, R. (2023). On the relation of instrumental dependence. In: N. Aussenac-Gilles, T. Hahmann, A. Galton, and M. M. Hedblom (Eds.) *Formal ontology in information systems: Proceedings of the 13th International Conference (FOIS 2023). Frontiers in artificial intelligence and its applications* (vol. 377, pp. 47–61). IOS Press.

Biccheri, L., Ferrario, R., & Porello, D. (2020). Needs and intentionality. In: B. Brodaric and F. Neuhaus (Eds.) *Formal ontology in information systems: Proceedings of the 11th International Conference (FOIS 2020). Frontiers in artificial intelligence and its applications* (vol. 330, pp. 125–139). IOS Press.

Blagec, K., Barbosa-Silva, A., Ott, S., & Samwald, M. (2022). A curated, ontology-based, large-scale knowledge graph of artificial intelligence tasks and benchmarks. *Scientific Data, 9*(1), 322.

Booch, G., Jacobson, I., & Rumbaugh, J. (1996). The unified modeling language. *Unix Review, 14*(13), 5.
Booch, G., Rumbaugh, J., & Jackobson, I. (1997). UML: Unified modeling language. Versão.
Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News, 538*(7623), 20.
Chen, M. (2021). Trust and trust-engineering in artificial intelligence research: Theory and praxis. *Philosophy and Technology, 34*(4), 1429–1447. https://doi.org/10.1007/s13347-021-00465-4
Chiffi, D., Moroni, S., & Zanetti, L. (2022). Types of technological innovation in the face of uncertainty. *Philosophy & Technology, 35*(4), 94. https://doi.org/10.1007/s13347-022-00587-3
Chung, L., Nixon, B. A., Yu, E., Mylopoulos, J., Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2000). The NFR framework in action. *Non-Functional Requirements in Software Engineering,* 15–45.
Cummins, R., Ariew, A., & Perlman, M. (2010). Neo-teleology. *Philosophy of biology: An anthology* (pp. 164–174). John Wiley & Sons.
Dalpiaz, F., Franch, X., & Horkoff, J. (2016). istar 2.0 language guide. arXiv preprint arXiv:1605.07767
Davis, A., Dieste, O., Hickey, A., Juristo, N., & Moreno, A. M. (2006). EffecDtiveness of requirements elicitation techniques: Empirical results derived from a systematic review. In: *14th IEEE international requirements engineering conference (RE'06)* (pp. 179–188). IEEE.
Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd international conference on machine learning* (pp. 233–240).
De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in Public Health, 11*, 1166120.
DeMarco, T. (2001). Structure analysis and system specification. In: Pioneers and their contributions to software engineering: Sd &m conference on software pioneers, Bonn, June 28/29, 2001, Original Historic Contributions (pp. 255–288). Springer.
De Souza, C. S. (2005). *The semiotic engineering of human-computer interaction*. MIT Press.
Dickover, M. E., McGowan, C. L., & Ross, D. T. (1977). Software design using: SADT. In: Proceedings of the 1977 annual conference (pp. 125–133).
Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. *41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210–0215). IEEE.
Eric, S., Giorgini, P., Maiden, N., & Mylopoulos, J. (2011). *Social modeling for requirements engineering*. MIT press.
Fensel, A., Akbar, Z., Toma, I., & Fensel, D. (2016). Bringing online visibility to hotels with schema.org and multi-channel communication. *Information and communication technologies in tourism* (pp. 3–16). Springer.
Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.
Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People–An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
Fumagalli, M., Bella, G., & Giunchiglia, F. (2019). Towards understanding classification and identification. *Pacific rim international conference on artificial intelligence* (pp. 71–84). Springer.
Fumagalli, M., Bella, G., Conti, S., & Giunchiglia, F. (2020). Ontology-driven cross-domain transfer learning. In: B. Brodaric and F. Neuhaus (Eds.) *Formal ontology in information systems: Proceedings of the 11th International Conference (FOIS 2020). Frontiers in artificial intelligence and its applications* (vol. 330, pp. 249–263). IOS Press.
Fumagalli, M., & Ferrario, R. (2019). Representation of concepts in AI: Towards a teleological explanation. In: A. Barton, Seppälä, S., and D. Porello (Eds.) *JOWO 2019—The joint ontology workshops, proceedings of the joint ontology workshops 2019, Episode 5: The Styrian autumn of ontology* (vol. 2518). CEUR Workshop Proceedings.
Giunchiglia, F., & Fumagalli, M. (2016). Concepts as (recognition) abilities. In: R. Ferrario and W. Kuhn (Eds.) *Formal ontology in information systems: Proceedings of the 9th International Conference (FOIS 2016). Frontiers in artificial intelligence and applications* (vol. 283, pp. 153–166). IOS Press.

Giunchiglia, F., & Fumagalli, M. (2020). Entity type recognition—dealing with the diversity of knowledge. In: *Proceedings of the international conference on principles of knowledge representation and reasoning* (vol. 17, pp. 414–423).

Griffiths, P. E. (1993). Functional analysis and proper functions. *The British Journal for the Philosophy of Science, 44*(3), 409–422.

Guarino, N., Guizzardi, G., & Mylopoulos, J. (2020). On the philosophical foundations of conceptual models. *Information Modelling and Knowledge Bases, 31*(321), 1.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? *Handbook on ontologies* (pp. 1–17). Springer.

Guha, R. V. (2011). Schema. Org support for job postings. Retrieved from http://blog.schema.org/2011/11/schemaorg-support-for-job-postings.html

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR), 51*(5), 1–42.

Guizzardi, G. (2005, Oct). *Ontological foundations for structural conceptual models*. Phd thesis, Research UT, Graduation UT, University of Twente.

Guizzardi, R., Amaral, G., Guizzardi, G., & Mylopoulos J. (2023). An ontology-based approach to engineering ethicality requirements. *Software and Systems Modeling 22(6),* 1897–1923.

Guizzardi, R. S., Guizzardi, G., Almeida, J. P. A., & Cardoso, E. C. (2010). *Bridging the gap between goals, agents and business processes*. iStar (pp. 46–51).

Guizzardi, R., & Reis, A. N. (2015). A method to align goals and business processes. *International conference on conceptual modeling* (pp. 79–93). Springer.

Hernich, A., Lutz, C., Ozaki, A., & Wolter, F. (2015). Schema. org as a description logic. In: *Twenty-Fourth international joint conference on artificial intelligence*.

Hevner, A.R., March, S.T., Park, J., & Ram, S. (2004) Design science in information systems research. *MIS Quarterly*, 75–105.

Horkoff, J., Aydemir, F. B., Cardoso, E., Li, T., Maté, A., Paja, E., Salnitri, M., Piras, L., Mylopoulos, J., & Giorgini, P. (2019). Goal-oriented requirements engineering: An extended systematic mapping study. *Requirements Eng., 24*, 133–160.

Hou, Y., Xiong, D., Jiang, T., Song, L., & Wang, Q. (2019). Social media addiction: Its impact, mediation, and intervention. Cyberpsychology: *Journal of Psychosocial Research on Cyberspace, 13*(1).

Jiang, Z., Xu, F. F., Araki, J., & Neubig, G. (2020). How can we know what language models know? *Transactions of the Association for Computational Linguistics, 8*, 423–438.

Kärle, E., Fensel, A., Toma, I., & Fensel, D. (2024). Schema. Org usage for hotels. Retrieved from https://schema.org/docs/hotels.html

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. MIT Press.

Kroes, P. (2010). Theories of technical functions: Function ascriptions vs. function assignments, part 1–2. *Design Issues, 26*(4), 85–93. MIT Press.

Lee, E. A. (2020). *The coevolution: The entwined futures of humans and machines*. Mit Press, 2020.

Lieto, A. (2022). The minimal cognitive grid: A tool to rank the explanatory status of cognitive artificial systems. *18th Annual Conference of the Italian Association of Cognitive Sciences (AISC)*.

Macdonald, G., & Papineau, D. (2006). *Teleosemantics*. Oxford University Press.

McDowall, R. (1991). The systems development life cycle. *Chemometrics and Intelligent Laboratory Systems, 13*(2), 121–133.

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems, 31*(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science, 56*(2), 288–302.

Millikan, R. G. (2005). *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press.

Mustafaraj, E., & Metaxas, P. T. (2017). The fake news spreading plague: Was it preventable?. In Proceedings of the 2017 ACM on web science conference (pp. 235–239).

Neander, K. (1991). The teleological notion of 'function'. *Australasian Journal of Philosophy, 69*(4), 454–468. Taylor & Francis.

Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2023). Taking AI risks seriously: A new assessment model for the AI act. *AI & Society,* 1–5.

Panasiuk, O., Akbar, Z., Şimşek, U., & Fensel, D. (2018). Enabling conversational tourism assistants through schema.org mapping. *European semantic web conference* (pp. 137–141). Springer.

Pascal, B. (1888). *The thoughts of Blaise pascal*. K. Paul, Trench & Company.

Peeters, R., Primpeli, A., Wichtlhuber, B., & Bizer, C. (2010). Using schema. org annotations for training and maintaining product matchers. In: *Proceedings of the 10th international conference on web intelligence, mining and semantics* (pp. 195–204).

Petrosino, S. (1999). *L'esperienza della Parola: Testo, Moralità e Scrittura* (Vol. 6). Vita e pensiero.

Porello, D., Bottazzi, E., & Ferrario, R. (2014). The ontology of group agency. In: P. Garbacz and O. Kutz (Eds.) *Formal ontology in information systems: Proceedings of the 8th international conference (FOIS 2014). Frontiers in artificial intelligence and its applications* (vol. 267, pp. 183–196). IOS Press.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd., GBR.

Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012). Semantic Cosine Similarity. In: *The 7th International Student Conference on Advanced Science and Technology (ICAST)* (vol. 4).

Robertson, S. E., Kanoulas, E., & Yilmaz, E. (2010). Extending average precision to graded relevance judgments. In: *Proceedings of the 3rd international ACM SIGIR conference on research and development in information retrieval* (pp. 603–610).

Saramago, J. (1999). *Blindness*. Houghton Mifflin Harcourt.

Shin, D., Kee, K. F., & Shin, E. Y. (2022). Algorithm awareness: Why user awareness is critical for personal privacy in the adoption of algorithmic platforms? *International Journal of Information Management, 65*(C), 102494.

Simon, H. A. (2019). *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird*. MIT press.

Sulé, A. (2015). Schema.org: An enhanced display of search engine results and much more. *Bid-textos universitaris de biblioteconomia i documentació* (34).

Tilley, S. (2019). *Systems analysis and design*. Cengage Learning.

Tort, A., & Olivé, A. (2014). A computer-guided approach to website schema.org design. *International conference on conceptual modeling* (pp. 28–42). Springer.

Van Lamsweerde, A. (2001). Goal-oriented requirements engineering: A guided tour. In: *Proceedings Fifth IEEE international symposium on requirements engineering* (pp. 249–262). IEEE

Von Frisch, K. (2014). *Bees: Their vision, chemical senses, and language*. Cornell University Press.

Wasson, C. S. (2005). *System analysis, design, and development: Concepts, principles, and practices* (Vol. 22). John Wiley & Sons.

Weber, R. A. (1998). *Information systems control and audit*. Pearson Education.

White, S. A. (2004). Introduction to BPMN. *IBM Cooperation, 2*(0), 0.

Wright, L. (1976). *Teleological explanations: An etiological analysis of goals and functions*. University of California Press.

Yu, E., & Mylopoulos, J. (1998). Why goal-oriented requirements engineering. In: Proceedings of the 4th international workshop on requirements engineering: Foundations of software quality (vol. 15, pp. 15–22).