THE BIMBOT: MEDIATING TECHNOLOGY FOR ENACTING COORDINATION IN TEAMWORK COLLABORATION

Ivan Mutis
Assistant Professor, Civil and Architectural Engineering, Director of the iConSenSe Laboratory,
Illinois Institute of Technology, Chicago, USA,
imutissi@iit.edu

Adithya Ramachandran
Graduate Student, Computer Science,
Illinois Institute of Technology, Chicago, USA,
ar1@hawk.iit.edu

SUMMARY: Today, BIM technologies in collaborative practice are widespread among construction project stakeholders. However, embracing either distributed or collocated tasks in collaborative practices is a complex, challenging activity. Each team member (actor) views collaborative design problems from a different 'lens', framed by the realities of their disciplines, experiences, and levels of engagement on tasks. The effect is a practice prone to conflict generation and misunderstandings among actors. BIM technologies and teamwork should be configured to adapt to one another in practice dynamically. The configuration should enable the effective performance of distributed and collocated work tasks. The presented study investigates these configurations to reveal constitutive aspects of how work should be executed in practice. The study focuses on adapting technology and teamwork to reveal a more effective way of delivering distributed and collocated work tasks. To explore the research question, two components were developed: a theoretical framework and a technology conceptualization. The framework presents fundamental constitutive elements in the coordination process. It illustrates the key aspects that draw the configurations of technology and teamwork. The technology concept is a design to assist in the execution of the tasks for coordination activities. It addresses the constitutive aspects of coordination for BIM processes in practice. The technology concept, named BIMbot, is a cognitive assistant that informs and advises on activities, engages team members together in a task, and facilitates fundamental actions for shared understandings, physical support, and informed advice. This paper contributes to shedding light on the difficulties for team members to reach a shared understanding of knowledge when they use BIM technologies. It presents the first development of the design of technology that provides actionable information to coordinate activities.

KEYWORDS: Cognitive Intelligent Agent, Coordination, BIM room, Cognitive Computing, Mediating Technology

1. INTRODUCTION

It has been nearly two decades since Building Information Modelling (BIM) reached a level of widespread use in the architecture, engineering and construction industry (Eastman et al., 2018, Mehrbod et al., 2019). It is recognized that the technology implementation is mature and has broadly impacted and transformed construction management practices (Demirkan et al., 2017). BIM has also empowered new forms of collaboration among stakeholders, thereby enabling greater integration between design and construction. Although BIM is often viewed as a set of tools and technologies as well as an essential technology intervention for collaboration (Dossick et al., 2014), it should also be accepted as a mechanism for mediating organization–i.e., an enabler for technology-based collaborative practices undertaken harmoniously and cohesively among participating actors.

New project delivery methods have evolved (by creating models of operation), accelerated (by launching and implementing innovative modes of operation), and scaled collaborative practices (through design operations in relevant teamwork). The delivery methods have incorporated BIM technology to engage disparate project stakeholders from different disciplines at earlier phases in the project (Dossick et al., 2014). These individuals are team members who manage separate work tasks that must be coordinated. Such collaboration relies heavily on technology and is referred to as collaborative technology practice (Orlikowski, 2007, Orlikowski, 1992). An example is the design-build method that encourages design teams to work collaboratively using technology for effective creation, knowledge exchange, and design-data management among members. BIM plays a crucial role in enabling the conditions for collaborative team practices, including those of knowledge sharing (e.g., shared understanding of designs) and technology operation on a shared computing platform (e.g., interoperability).

Participating actors from multiple disciplines – ranging from mechanical to structural to construction engineering to architecture – intervene across organizational boundaries within and (or) between firms in BIM collaborative practices. They engage in distributed and collocated, collaborative, and technology-ready tasks. The term 'distributed' refers to the actors' execution of project-related design and construction activities from various locations in the absence (both in real-time or at different points in time) of in-person interaction with other disciplines. To overcome the absences, individual actors may engage with one another remotely using a mediating technology when executing distributed work (Orlikowski and Scott, 2008). 'Collocated' refers to sharing workspaces – in the same physical location – to execute tasks with other team members to facilitate in-person interactions.

When embracing either distributed and collocated tasks, the challenge in BIM collaborative practices is that each team member (actor) views the collaborative design problem from a different 'lens', framed by the realities of their disciplines, experiences, and engagement levels on tasks. The effect is a practice prone to conflict generation and misunderstandings among actors. As an example, the heterogeneity, roles, and interests of team members can lead to conflicting views on solutions when interpreting engineering designs, which impact the effectiveness of the team task execution. Although BIM enables a more robust connection and technology-based collaborative practice among participating actors, engaging team members across organizational boundaries remains problematic. BIM technologies allow connections, but have yet to harmonize their members' actions, as their constitutive technology design does not aim to synchronize and unify disparate actions. Thus, there is a need to investigate how technology can best assist collaborative practices to harmonize team members' actions, including the consolidation of distributed and collocated work tasks.

BIM technologies and teamwork should be configured to adapt to one another in practice dynamically. The configuration should enable the effective performance of distributed and collocated work tasks. Research on these configurations may reveal constitutive aspects of how work is executed in practice. Research on adapting technology and teamwork may reveal a more effective way of delivering distributed and collocated work tasks. The expected outcome of new configurations is a new design of ongoing BIM processes – a design that accounts for enacted conditions and mechanisms that facilitate distributed and collocated work tasks. This research will facilitate understanding of how new configurations of technology and teamwork should pervade in practice by focusing on coordination as a mechanism for synchronizing and unifying actions to achieve harmonious operations. In addition, research on coordination tasks will shed light on what aspects of technology and teamwork should be configured to effectively execute distributed and collocated work.

The emphases of these questions depart from the view of mainstream research, which has focused primarily on BIM affordances for coordination (e.g., BIM technology capabilities and its tools for coordinating design team members) and on barriers (e.g., technologies that make the BIM process less effective in practice) to constitutive aspects of technology use (e.g., BIM and other technologies enactments for its use in coordinating tasks). Barriers

include enabling (i.e., new possibilities to connect and share information across specialties) and constraining (i.e., difficulties to engage in interpretation to design information flowing from different disciplines) collaboration.

To explore the new views proposed by the questions of this research, two components drive the study: a theoretical framework and a technology conceptualization. The theoretical framework will present fundamental constitutive elements in the coordination process. The technology conceptualization will design mechanisms to assist in the execution of tasks for the coordination of activities.

The theoretical framework aims to illustrate the key aspects that draw technology and teamwork configurations (see section 2). It serves to navigate answers to the research questions to facilitate solutions for enacting conditions and mechanisms for distributed and collocated work tasks using new technology. It permits the analysis of factors that have the power to establish or enact how coordination practices are materialized. It highlights the need to examine how those factors – as constitutive of technology use – might differ and under what circumstances. It helps identify critical new issues and describes the view of coordination, whose emphasis is on how coordinating tasks are taking place, or instantiated, in practice and how technology can contribute to such instantiation. Instantiations, herein, refer to actions to materialize in times and places while coordinating. The framework provides a common language and a frame of reference that emphasizes how BIM technologies might be used or configured to help accomplish the BIM process. It is worth noting that the BIM process may include not only coordination of data-rich, object-based parametric representations but also intersecting management activities ranging from project documentation to planning to scheduling.

The technology conceptualization addresses the constitutive aspects of coordination for BIM processes in practice (see section 3). The technology concept, named BIMbot, is a cognitive assistant that provides actionable information for coordination of activities; informs and advises on activities; engages team members together in a task; and facilitates fundamental actions for shared understandings, physical support, and informed advice. BIMbot is a conversational Question-Answering system that responds to any unseen query built on a deep learning-based sequence-to-sequence model. The full development and deployment of the technology concept is an enormous endeavor, including transformational aspects of the technology in the BIM process. This paper presents the first development of this technology concept, including its system architecture, challenges for implementation, and prototype testing. The initial architecture is based on the current state of the art in natural language processing techniques.

2. THEORETICAL FRAMEWORK

2.1 Coordination and Communicative Actions

Coordination among team members is a complex process of managing resources and activities. Coordination may occur at the same time and place and distributed and collocated in time and space to achieve adequate performance. It is the mechanism for synchronizing and unifying the actions (Okhuysen and Bechky, 2009) in distributed and collocated work, whose goal is to achieve harmonious operations (Malone and Crowston, 1990) in teamwork tasks (i.e., coordination facilitates the teamwork actions for distributed and collocated work). Coordination, herein, is a set of tasks and processes by which team members carrying out project activities manage interdependencies to perform effectively as a group (Crowston, 1997).

Multiple factors, including the nature of tasks, time of execution, type of mediating mechanism, and group characteristics, make coordination challenging and impact teams' performance. As a result, team members shape the activities of a group by performing routine adjustments and adaptations to achieve coordination. Adjustments and adaptations are essential routine tasks to enable conditions for knowledge sharing (Bechky, 2003). For example, setting the conditions for an efficient, shared understanding of designs to address, for instance, their unclear characterization (scope, priority, and rationale) (Mehrbod et al., 2019), is an adjustment and adaptation task using mediating technologies (BIM) in a common computing platform. However, implementing adjustments and adaptations is challenging, and their enactment is poorly understood (Beane and Orlikowski, 2015). Putting into action adjustments and adaptation make it possible for a team member to actively respond to unforeseen situations. Team members incorporate ad-hoc, improvised, and un-organized responses to adjust and adapt. Inefficiencies of ad-hoc, improvised and un-organized responses hinder the team member's effectiveness in responding to unforeseen situations. These inefficiencies lead to poor enabling conditions for knowledge sharing (Carlile, 2002).

Team members may incorporate norms for adjustments and adaptions as more effective methods to enable knowledge-sharing conditions. Norms are collective representations of acceptable operations of a group or individuals, and they are commonly understood as considerations of principles of right actions. Putting into action or enacting norms implies launching remedial actions to guide or control acceptable activities for coordination. Understanding norms is critical for coordination, especially when establishing and enacting distributed and collocated work (Ghosh, 2004). Putting into action or enacting norms within distributed and collocated work conditions have their own challenges and are a daunting endeavour. For example, when team members are distributed in time and space, team members engage in agreements remotely as opposed to in-person, which makes it challenging to reach agreements for actions.

Further, norms are more provisional settlements (Kaplan and Orlikowski, 2013) than shared considerations and understandings of principles of right actions. These characteristics add conditions of instability for longer time spans. Provisional settlements (Kaplan and Orlikowski, 2013) entail setting processes of calibration and exploration of solutions. They enable conditions for knowledge sharing by adjusting and adapting current tasks for remedial actions.

Team members use communicative actions to enact norms and to subsequently facilitate provisional settlements for right actions. Communicative actions establish prospects and alignments of goals to prevent undesirable problems (Ghosh et al., 2004) among team members. For example, using communicative actions to execute provisional settlements to execute team tasks, team members align local conditions as a team task goal, such as local regulation in a project. Communicative actions are the enabling mechanisms that permit team members to respond to problems and engage in remedial actions. For example, using communicative actions, team members participate in the exchange of information on design changes – either in real time or dispersed in time – to prevent the delivery of incorrect BIM data. The exchange arrangements are provisional and achieved settlements through the exploration of solutions among team members. The effectiveness of communication (communicative actions) becomes critical for desired responses and achievements of stated objectives and enabling knowledge sharing conditions. Communicative actions play a critical role in coordination, not only for routine adjustment and adaptation but also for distributed and collocated work over a period of time.

Communicative actions have linguistic, representational, and structural forms. The recognition of these forms and their relevance (Orlikowski and Yates, 1994) for team members is critical (i.e., recognition and relevance lead to the communicative action). Team members invoke forms when they require and recognize purposes to discuss, make decisions, and respond to requests from other actors. The forms have important implications in technology design. For example, while actors embed themselves in the BIM process using linguistic, representational, and structural forms, characterization of the communicative actions from the lens of these forms would help understand technology design to assist the BIM process. Features of these forms follow.

Representational forms are mediums. They allow participants to enact communicative actions. Examples of these forms are chatrooms. Chatrooms are a medium where discussions are invoked with specific goals for decision making (e.g., virtual message exchange mechanisms to convey observable aspects or views of BIM designs). Linguistic forms are text, speech acts, or utterances. Linguistic forms serve to exchange messages using domain or common vocabulary. Linguistic forms may be embedded in representational forms (e.g., chatrooms). Structural forms are feature elements that obey an organization or arrangement. Examples are hierarchies in teams and arrangements among stakeholders (e.g., having team member leaders, building requests for participation in a virtual weekly meeting).

2.2 Communicative action in BIM processes: communicative act

Communicative actions engage the fieldwork of engineers, architects, contractors, and all associated team members in BIM processes, such as the actions in the BIM room (Merschbrock and Munkvold, 2015). These actions take place in collaborative virtual or in-person meetings, and they engage the linguistic, representational, and structural forms.

Viewing BIM as a mediating mechanism in the BIM processes, communicative actions reveal two essential relationships: (1) between the actor (a group of actors), who is the source or origin of the representation and the mediating representations and (2) between the actor (a group of actors) as interpreters and the mediating representation. The relationships in BIM processes are among actors and established goals of messages.

Communicative actions in the BIM process have a specific goal and imply a social recognition of forms. A specific goal denotes the distinctive indication of specific actions. For example, in sharing information, the goal is to align

interpretations when not all geometrical content is represented on BIM representations, which is a case when there are not confirmed as-built conditions or when there are lower levels of development specification (LOD) content in the design phase. In the BIM process, actions that incorporate norms that govern the activity result from exchanging, sharing, and interpreting information.

The unit of observation, namely the *communicative act*, is proposed as the basic framing entity to facilitate the understanding of communicative actions and their role in the BIM process. The *communicative act* frames the actions that take place and the method when actors work together to complete activities pertinent to a goal. For example, a *communicative act* informs how a set of actions from the source to the interpreter occurs, and vice-versa, to produce effective communication in the BIM process. Further elaboration on the method to use the *communicative act* and analyze communicative actions and their role in BIM process follow.

2.2.1 Speech acts

The authors drew from Searle's speech act theory (Searle, 1969) the analysis of the linguistic forms of the unit of observation (*communicative act*). The speech act theory posits that a speech act defines rule-governed forms of behavior when actors play a role in communicating information through language, specifically through the act of speaking. The theory suggests the existence of sufficient and necessary conditions for the performance of particular speech acts where certain kinds of behavior can be characterized (e.g., intentional behavior). Speech acts treat linguistic forms through the act of speaking. The governing rules and conditions from speech acts serve as indicators to frame the linguistic forms in *communicative acts*.

The *communicative act* has a broader conceptualization than the speech act. The *communicative act* includes representational and structural forms during any act of speaking, which leads to a wider discourse context.

The speech act characterizes linguistic forms using the act of speaking. Speech acts emphasize what the speaker communicates to the hearer by relying on the mutually shared background of the information or contexts and the intention of the utterance (Searle, 1995, Searle, 1969). In the simplest case, two actors– a speaker and a hearer– participate in a speech act. Thus, when there is an utterance, an understanding of the facts and relevance of the conversation, a setting up of the background information pertinent to the conversation, including assumptions and inferences, are needed to capture the intended meaning within the expressed utterance.

Actors in the BIM process engage linguistic forms as in the speech acts to capture the intended meaning. In this process, actors also engage representational and structural forms with a particular goal (e.g., exchanging, sharing, and interpreting information). For example, when interpreting information in a *communicative act*, actors exchange utterances on a shared representation (e.g., BIM). One or more actors from a team generate the information, and others are the interpreters. In the speech act, there is the speaker and the hearer. In the *communicative act*, there are interpreters (receivers) and sources who are participating members of the team using a medium (i.e., chatroom, virtual room). Each team member has a predefined role. They share BIM designs as representations of concepts to other team members – a structural form of communicative action. In this action, different forms of representation (e.g., BIM or linguistics representations) mediate to communicate meanings to actors and suggest some actions for the interpreter – a representational form of the communicative action. Then, the interpreter makes assumptions and inferences to enact the intended meanings. Thus, the interpreters identify the intention of the representation from the source and assert their meanings – a goal of the *communicative act* and a parallel to the speech act. The outcome from asserting is the interpreters' commitment to actions on a particular activity. If the assertion does not capture the meaning of the representation, the interpreter directs a request for additional information to the sources – through requests and assertions or speech acts.

From Searle's work, the hearer asserts and commits to action or rejects the utterance in play when the intention was not captured within the speech act. Speech acts can further be broken down into expressions of force. The theory defines three types: locutionary, illocutionary, and perlocutionary. Locutionary acts are utterances (locutions) with certain meanings. Illocutionary denotes carrying a directive for the audience (e.g., promising, ordering, apologizing). An illocutionary act is the entire speech act. An utterance materializes the illocutionary act (i.e., illocutionary acts can be called illocutionary utterances). Illocutionary sometimes contains elements of reference, predicate, and force. Perlocutionary brings the consequences to the audiences (e.g., showing the effect of the locution – inspiring, persuading, or deterring).

Technology design for communicative actions in the BIM process should include illocutionary and perlocutionary acts, as it would enable *communicative acts* (i.e., the actions that take place and how actors work together to complete activities pertinent to a goal).

Communicative acts incorporate forms that enable a wider discourse context rather than chains of independent illocutionary forces. *Communicative acts* create an observable unit in the BIM process to yield more efficient activities that required situated dialogues among team members, either in-person or in virtual environments.

3. ASSISTING COMMUNICATIVE ACTIONS USING TECHNOLOGY: COGNITIVE ASSISTANTS

The presented technology concept aims to harness communicative actions for coordination for BIM processes. The design pivots on enabling *communicative acts* to ease BIM processes. Actors or team members interact with a dialogue-based agent named BIMbot to generate human-collaborative processes using a question and answer (Q-A) system. The system's goal is to support actors to enhance their capabilities in associating and retrieving information required in coordinating activities in a wide variety of tasks. Actors need to make decisions and take actions to strengthen effectiveness by reducing their proneness to errors and misjudgments. While the effectiveness of some Q-A systems is comparable to humans, the approach is not designed to include comprehensive forms of social dialogue but to incorporate *communicative acts* that intervene in the BIM process.

The dialog-based agent design was based on machine learning and deep learning algorithms. Taking advantage of today's machines' higher computing power, the agent design adopted a cognitive computing approach. According to (Coronado et al., 2018), *cognitive computing* refers to intelligent systems that learn at scale, reason with purpose, and interact with humans and other systems naturally (Demirkan et al., 2017). Figure 1 shows cognitive computing's evolution process (adapted from Chen et al. (2018)). The cognitive assistant builds on deep learning approaches (neural network-based computational models), composed of multiple processing layers to learn representations of data with various levels of abstraction (LeCun et al., 2015). With improving computing capabilities for training using parallel computing, hardware accelerators, and machine learning packages, deep learning has become an ideal way to process natural language and build cognitive computing applications.

3.1 Review of cognitive computing agents and applications

The current suite of tools and services for cognitive computing with different interfaces are mainly available for voice, text, and vision. Popular voice-based conversational systems include several commercial products, such as Amazon's Alexa™, Apple's Siri™, and Microsoft's Cortana™. There are many toolkits available that deal with programming and customizing these platforms, including and not limited to Google's dialog flow™, Amazon Poly™, Amazon Lex™, Google's Vision API™, and Amazon Recognition™, to name several. Early popular development was IBM Watson™. This agent is a type of cognitive computing system that provides services as Question-Answering. It processes natural language with high capabilities with different functionalities (e.g., interacting with users, analyzing structured and unstructured content, making connections and drawing relationships between data, ontology analysis, and clinical trial matching to find potential clinical traits). The use of cognitive assistants is evolving and ubiquitous at different levels of automation features with different interactive designs. Examples lay in many domains with different interactive designs ranging from augmentation of data-driven decision-making cognitive support systems for automation of labor-intensive tasks in accounting and auditing (Li and Vasarhelyi, 2018) to applications in building analytics for the execution of business processes (Zasada, 2019), reasoning tasks in learning (Le and Wartschinski, 2018), to healthcare (Preum et al., 2021), among others.

The cognitive assistants are emerging decision support tools designed to augment human capabilities by providing contextual knowledge using dialogs and gestures. The evolution of these tools aims to enhance cognition and intelligence (Siddike et al., 2018), using collaborating forms to arrive at better performance in activities of interest. The cognitive agent serves the user, thereby creating sustainable value that benefits the agent's system–e.g., collecting data to refine its model– and the user–by servicing the activity of interest by proving cognition power. Cognition power has multiple dimensions such as reasoning (revealing meanings), validating results (complying), retrieving data (searching, recognizing patterns), analyzing (identifying similarities and differences). Cognitive assistants design includes learning and processing capabilities to predict for assisting in decision-making and executing a task or process.
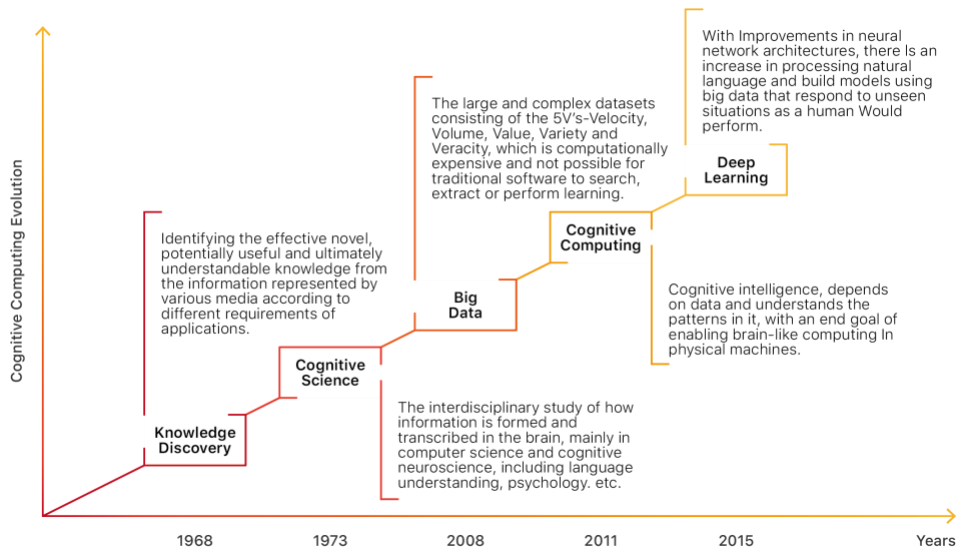
FIG. 1: Cognitive Computing Evolution, Adapted from Chen et al. (2018)

Other approaches in the project management arena using cognitive computing include a personal assistant for relieving one's self from managing g and scheduling tasks, i.e., PExA™ (Project Execution Assistant). PexA has been developed to improve a knowledge worker's productivity and effectiveness by aiding in organizing and performing tasks (Myers et al., 2007) for time and task management in an individual's work. The entire system components interact using an asynchronous messaging scheme to exchange data, queries, and requests for performing tasks, resulting in a distributed and modular environment that simplified overall system design.

There are many commercial applications, mostly for customer service using dialog features (e.g., chatbots), to reduce human labor with an automated response system. Advantages are rapid responses to a customer query, fewer human errors, and informational responses, among others. Approaches in commercial applications have been developed by Xu et al. (2017) using deep learning and natural language processing.

3.2 Cognitive Agent – BIMbot's Architecture

The introduction of Machine Learning (ML) paved the way for widespread creation of intelligent conversational agents. These agents are called intelligent as they can learn from data or previous dialogs. Two types of models use ML and exploit its advantages: (1) retrieval-based models and (2) generative models. Retrieval-based models do not create any new responses. These models generally select an answer from a pool based on the question received. A simple approach can be viewed as querying a database for a search with no emphasis on the semantic understanding of the input. The searches are unsuccessful in language processing tasks, especially Q-A systems, due to their inability to incorporate context information for queries.

On the other hand, the design of generative models aims to understand a user's input. Generative models can create their own answers based on the question they receive, and they learn the structure of sentences through training. Generative models require a significant amount of data for training purposes. These models outperform retrieval-based models after training, particularly for new queries presented by the user (Cahn, 2017). Currently, the majority of research on generative models uses deep learning techniques. Deep learning architectures, e.g., sequence-to-sequence, can efficiently process text. Based on generative models' success, the presented cognitive agent's architecture for the BIM process – the BIMbot cognitive agent – uses a generative model along with a rule-based model that runs on pattern matching [10,11]. BIMbot will use a combination of technologies and resources for a successful pro-active exchange of stakeholder dialogs. The upcoming sections will elaborate on four major components of BIMbot's architecture (see FIG. 2): 1) Corpus; 2) Neural Machine Translation (NMT); 3) a rule-based model; and 4) NLP Engine.
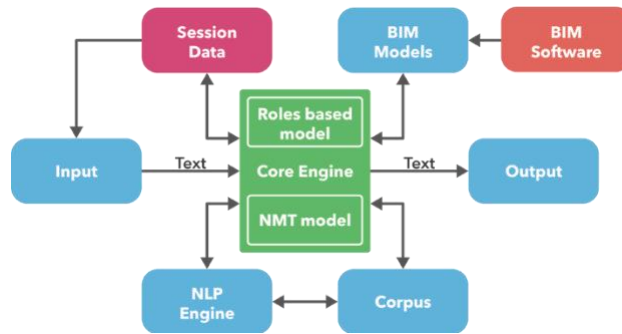
FIG. 2: BIMBOT's Architecture

3.2.1 Corpus

One of the agent's most important components – to denote domain knowledge and general domain knowledge – is the corpus, which is represented and stored as a collection of texts (data) in an unstructured format. The corpus is pivotal for the system architecture as it is the source of data for training. There are two steps involved in corpus creation. The first consists of data collection as input for the corpus and the second is cleaning the collected data.

The first step was to source data for training purposes, using the following general domain knowledge and specific domain sources: 1) Cornell Movie Corpus, 2) Reddit, and 3) conversations in BIM rooms. Data was collected from posts (approx. 3848330 posts) and speech acts (set of 10 minutes to 30 minutes videos, using automatic caption technology for text conversion). The combined data size for training was approximately 80 GB. Data was aggregated from three sources to have both open-domain knowledge (general knowledge) and specific domain knowledge (BIM-related knowledge).

The collected data, which is usually unstructured, was processed for cleaning in the second step before employing it as input for training. The cleaning process consisted of two steps. The first was to structure the sentences in Question and Answer pair format, as described in Table 1. For this purpose, separate python scripts were written for the Reddit dataset and Cornell Movie Corpus datasets. Following this step, the researchers identified semantically colorful words in the structured data (i.e., terms that would be irreverent to teamwork conversations) to truncate the Question and Answer pair. This process allows modeling the cognitive agent responses that included features of politeness and patience, among others.

The researchers used two formats (text and audio) for processing information from the BIM-room conversations. Examples of tools used were IBM Watson™'s speech-to-text web application. The resulting data from audio were aggregated with existing data in text format to structure the Question and Answer format. The *NPL module* managed and analyzed data input from the worker. Input methods were via manual typing or speech-to-text (e.g., using open source toolkits such as CMU Sphinx (CMUSphinx, 2017)). Pre-processing techniques (tokenization and segmentation, noise removal, normalization) were used to reduce useless words, syntax errors, and disfluencies (fill-pauses, speech repairs, corrections, repetitions) in the utterances. The output of the analysis included, for instance, syntactic structures that were used as input in the semantic interpreter module.

Table 1. Cleaned data – Question and Answer format

Can I go out to play?	Are you done with work?
I cannot ignore the verdict of my council	Surely you can do anything you want.

3.2.2 Neural Machine Translation - Generative Model

Neural Machine Translation (NMT) falls under the umbrella of sequence-to-sequence models (Wu et al., 2016, Sutskever et al., 2014), and is also known as encoder-decoder architecture. The BIMbot cognitive agent model used a Recurrent Neural Network (RNN) (Cho et al., 2014) approach, as it is commonly implemented when working with encoder and decoder models.

The rationale is based on the human tendency to persistently pivot on the same pattern of assertions when it comes to decision making (i.e., when humans make decisions, it is usually based on their previous thoughts retrieved from long-term memory). Simple Deep Neural Networks are widely used for non-sequence datasets, such as with fixed dimensionality vectors. But they perform poorly for time series data or sequence-related data as used in

Natural Language Processing, which is a significant limitation as many important problems are best expressed with sequences whose lengths are not known a-priori. Question-answer can be seen as mapping a series (sequences) of words representing the question to a series (sequences) of words representing the answer (Brownlee, 2017, Sutskever et al., 2014). Thus, conventional neural networks don't function as humans do, and this is a significant drawback. With the introduction of Recurrent Neural Networks (RNN), this issue is minimized. RNNs have loops enabling them to hold information in memory.

RNN models vary in terms of (1) "directionality – unidirectional or bidirectional"; (2) "depth – single or multi-layer"; and (3) "type - a Long Short-term Memory (LSTM), or a gated recurrent unit (GRU)". For current research that demands to keep long-term dependencies in memory, the researchers opted for an RNN model, which is bi-directional, multi-layered, and has LSTM units. Long-term dependencies are important since the context (conversation) between actors and the cognitive agent increases. The agent should incorporate the context to generate appropriate responses (Zachary C. Lipton, 2015, Karpathy, 2015).

There are two phases involved to decompose the learning process (see FIG. 3): (1) encoder and (2) decoder. The NMT is an approach that spreads the source sentence with an encoder's help to build a thought vector, as explained in (Luong et al., 2017). The thought vector is a sequence of numbers that represents the sentence meaning. A decoder then processes this thought vector to produce the correct response.
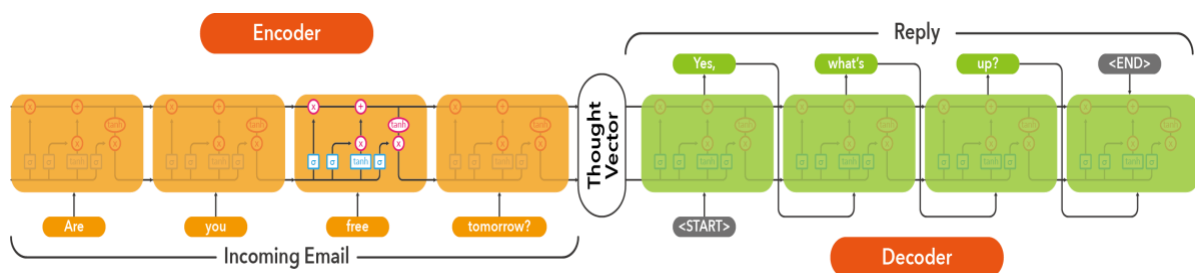

FIG. 3: Sequence to Sequence Model

In the encoder phase (FIG. 4), the source sentence from the corpus is sent to the encoder to generate the source context using the present word and the previous source context. The "source context" represents the words in numbers. Each word has a source context associated with a series of numbers, as illustrated in Fig. 4. The representation of numbers to words is done using an embedding layer. The embedding layer was built on words from the general domain and AEC vocabulary files across higher dimensions. Words that have similar and common properties had the same dimension of space. For example, parts of speech were found on the same dimension, while gender was on another dimension. The basic approach of word embedding was using vector space models to represent words in a continuous vector space where semantically similar words were mapped to nearby points. Global Vector (GloVe) is an unsupervised learning algorithm that was used to obtain the vector representations for words (Addor and Santos, 2014). The approach provides high accuracy for arriving at a semantic understanding of the corpus and can improve the cognitive agent's performance (Pennington et al., 2014).
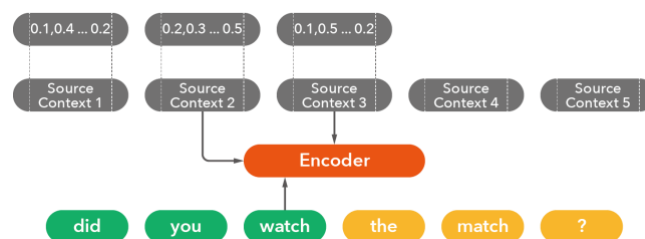
FIG. 4: Encoder Phase

The target words and responses were generated in the second phase. The approach uses target context, source context, and the previously generated word (FIG. 5). The target context contains the status of text generation, while the source context, along with the attention model (Bahdanau et al., 2015), includes the representation of the source sentence. The attention model calculates attention weights, as described in (Bahdanau et al., 2015), for the source context. This representation is called memory. At each time step, the target context was used to select which part

of the memory must be read. The approach enables the NMT model to focus on the required memory to create an appropriate response. The text generation stops when the decoder uses an end-of-statement tag (Sennrich et al., 2016).

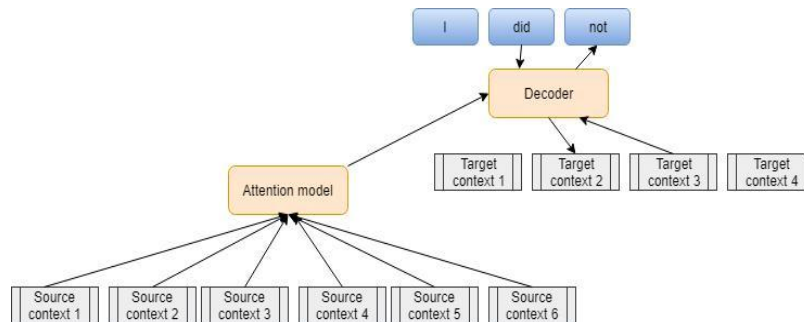*FIG. 5: Decoder Phase*

3.2.3 Rule-based model

The approach designed a rule-based model to search for additional information from the actors' input query when requiring domain and project-specific information. The researchers used closed domain information from the project-specific BIM to offer features like searching. The authors parsed BIM data from IFC files and created an ifcXML file to search and extract text associated with the designs (e.g., properties, dimensions related to a BIM model) from ifcXML files.

3.2.4 Natural Language Processing (NLP) Engine

The NLP engine forms the main text processing module in the system architecture. Natural language consists of text in sequential order. The text is broken into individual tokens that form a single lexical unit. The tokenizer handles the lexical unit, which splits the texts (data) from the corpus into words (Bird et al., 2009). The tokenizer helps create the vocabulary file used by the system during its learning (inference) phase. A vocabulary file consists of all word types in the corpus of documents, and it is used to form a closed domain vocabulary. An additional feature to this includes a <unk> token (to indicate the unknown word or missing words) with which it is assigned for any new occurrence of the word in the text.

4. IMPLEMENTATION AND DEMONSTRATION

There were two phases in our experimental setup. The first phase is training using the corpus data, and the second is the inference to test how the cognitive agent responds to unseen input.

To train the cognitive agent, the authors gathered around 300,000 pairs of conversations. For the cognitive agent to have a reasonable learning rate, it was necessary to provide the correct number of epochs. An epoch is a forward pass and backward pass of all training data. Additional training would be required to increase the number of epochs. It is also important to consider overfitting or underfitting the data when having a high and low number of epochs. The researchers trained our cognitive agent for 36 hours on an Nvidia P5000 GPU to create the base model. For the cognitive agent to learn diverse topics in the general domain, it was necessary to provide significant volumes of data (the implementation data set used approximately 80 Gigabytes). For the experimentation, after training, the researchers were able to get responses to unseen source sentences. This process is called inference. The cognitive agent was needed to find the semantics of the sentence to gauge if the sentences were general conversations or about the BIM process (domain-specific). The cognitive agent was able to find the semantics using the rule-based model for the domain-specific and the NMT model for the general domain by responding to the user queries, as shown in FIGs 6 and 7. A manual (user's) input was required to make the agent switch between the general-domain model (NMT) and the domain-specific (rule-based) model, per the user's intent of the query.

For evaluation, the researchers implemented a manual approach considering two metrics: user engagement and speed. Standard metrics like precision and recall quantifying the performance of text generation tasks were not suitable for this type of system due to the difficulty of factor responses (multiple ways to approach answers) to a particular question.

Team member (user)	Cognitive agent (BIMbot)
Do you have the law office door schedule?	Sure, it is open. Do you need anything from it?
What is the area of 100 A?	500 SF
What is the floor finish of 100 A?	Covet CT

Domain-specific communicative acts

```
Hi. Do you have the law office door schedule ?
> Sure, It's open. Do you need anything from it ?
What's the area of 100A ?
> 560 SF
What's the floor finish of 100 A ?
> CT-1
What's the base finish of 102 ?
> Coved CT
What's the wall finish of 106 ?
> VWC-1
Do you also have the law office room schedule ?
> Sure, It's open. Do you need anything from it ?
Can I have the width of mark 101 ?
> 3'-4'
What about the length of mark 106 ?
> 6'-0'
Thanks
>
```
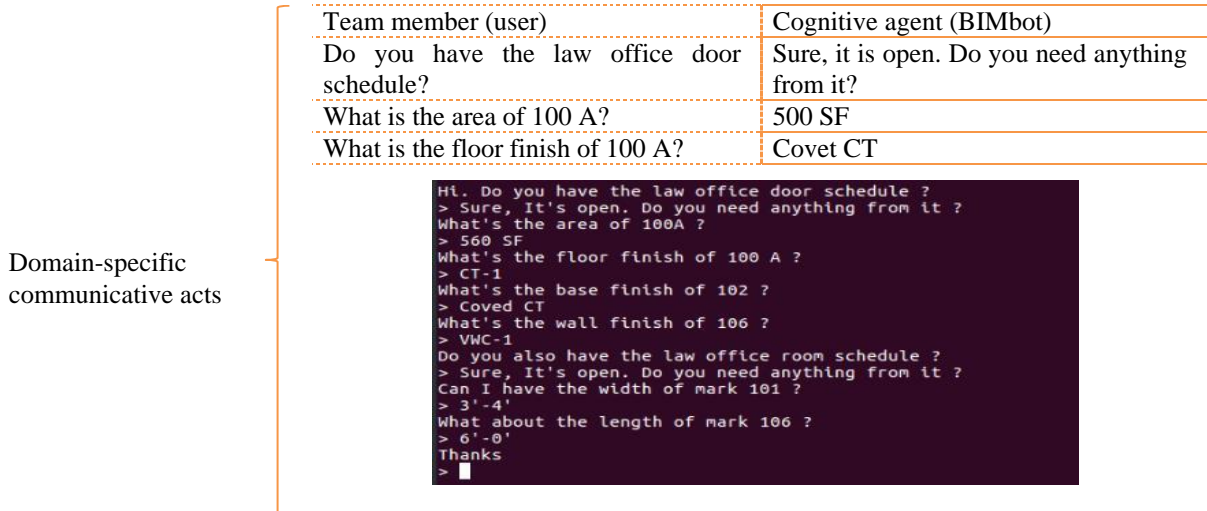
FIG. 6: Communicative acts: requesting domain-specific information.

User engagement is a qualitative assessment in which the cognitive assistant starts conversations and responds to users' requests by keeping the conversations coherent. Graduate students evaluated the dialogs' consistency, giving yes or no answers that seem the most reasonable or logical. BIMbot could continue the conversation coherently up to four questions and responses. The framework was a flow: (1) start, (2) give context to enable open questions, (3) ask a question; (4) if answers reasonable or logical continue question (step 3), if not (step 5); (5) rejects the logic and stop the flow. When the 5th question and response were provided on average, the reasonable and logical answers were rejected- i.e., the system rendered an answer that does not relate to the question. A possible counter to this issue could be changing the NMT's model parameters, as discussed in Pennington et al. (2014).

Speed refers to the qualitative estimation of time when getting the answer from the user query. The developed cognitive assistant was able to respond to a user's request in a fraction to a few seconds (0.5-2s).
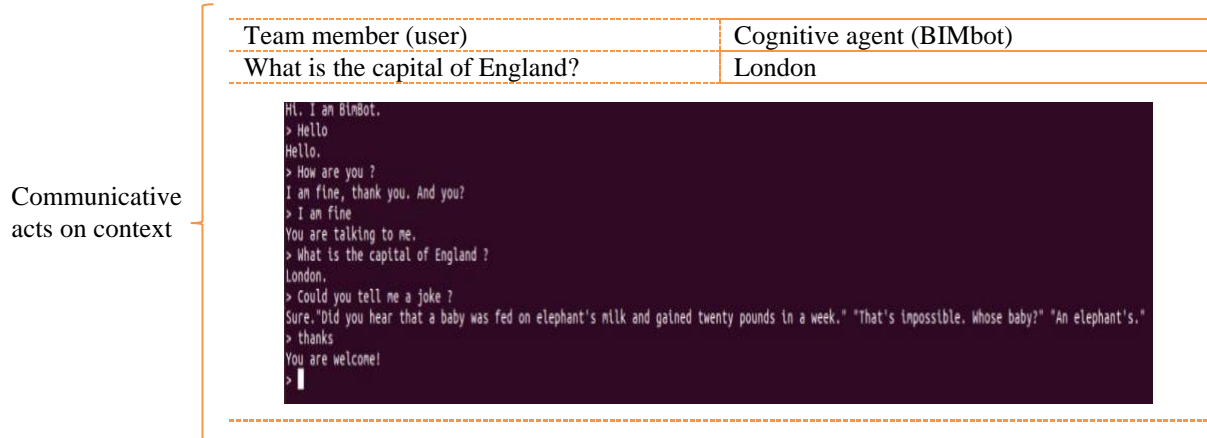
Team member (user)	Cognitive agent (BIMbot)
What is the capital of England?	London

Communicative acts on context

```
Hi. I am BimBot.
> Hello
Hello.
> How are you ?
I am fine, thank you. And you?
> I am fine
You are talking to me.
> What is the capital of England ?
London.
> Could you tell me a joke ?
Sure."Did you hear that a baby was fed on elephant's milk and gained twenty pounds in a week." "That's impossible. Whose baby?" "An elephant's."
> thanks
You are welcome!
>
```

FIG. 7: Communicative act: requesting context-related information.

4.1 BIMbot dialogs and coordination

The difficulties for team members to reach a shared understanding of knowledge in the BIM process demands key routine adjustments and adaptations in activities at hand – or adjustments and adaptation for coordination activities. The adjustment and the adaptation enable conditions for knowledge sharing in the BIM process. BIM is a mediating mechanism to facilitate adjustments and adaptations. BIMbot is a technology mediation that uses a dialogical approach to interact between the team member and the cognitive agent using communicative acts. See, for example, dialogs from Figure 6, where the user subsequently requests the agent about current information needs. The BIMbot design pivots on enabling *communicative acts* to ease BIM processes to facilitate team members' enactments for coordination activities.

The agent's answers facilitate adapting current conditions such as lack of information at hand, lack of awareness of contextual information, and others. There are multiple scenarios where the agent's intervention might be useful, along with all phases of the project, when critical information sharing tasks are paramount. For example, during the coordination of designs from multiple stakeholders, project participants actively require associated information of definitions of a design component, which are yet to be fully specified (e.g., arriving at common naming definition or conventions such as requirements, material quality, equipment specifications). The agent can provide predictions of definitions after a set of questions and answers using question and answer formats–the agent predicts information for the users' interpretation after queries, based on a dialogical approach (question and answers). Answering complex questions requires multiple reasoning steps, which gradually adjusts the region of interest to the given image's most relevant part. It would be an agile method to retrieve meaningful information from different sources to facilitate its association with the users.

Figure 7 shows an example of the *communicative acts* when information of context is required. The adoption and the adjustment reduce efforts to search for information to other team members, including contextual issues that impact the BIM model to other team members. Thus, the strategy addresses how team members can adjust and adapt to better conditions by facilitating enactments using BIMbot agents in BIM processes. It is expected that the agent will reduce ad-hoc, improvised, and un-organized responses to counteract to unforeseen situations while team members execute coordinating activities. Using a cognitive agent is a technological approach that enables distributing work – actions of coordination – since coordination is both a challenging endeavor and a difficult goal. The dialogical approach facilitates the sharing of knowledge among individual actors. The dialogical approach aims to reduce barriers from different views of realities from their disciplines, experiences, and engagement levels on tasks. The employment of *communicative acts* should reduce conflicts and misunderstandings routinely embedded in collaborative tasks.

5. CONCLUSION AND FUTURE WORK

The presented approach addresses challenges of communicative actions for coordination by designing a mediating technology that delivers general and specific domain information following a dialogical approach (question-and-answer system) among team members. The dialog facilitates stakeholders' connections, learning, and reducing barriers by providing and adapting synthesized information for use from different contexts (general and specific domain context). The technology is a cognitive agent. Its operations are framed in *communicative acts* to represent actions that take place when actors work together to complete activities pertinent to a goal in a BIM process (e.g., actors' question and agent's answer to retrieve properties of a BIM design with ease). *Communicative acts* inform team members on issues that may arise in the BIM process, thereby facilitating knowledge sharing among team members. When using *communicative acts*, the cognitive assistance serves as a form of a knowledge broker when coordinating tasks.

The presented research built a theoretical framework to introduce a conceptualization of communicative actions. It demonstrated how communicative actions play a critical role in coordination not only for routine adjustment and adaptation, which enable conditions for knowledge sharing but also for distributed and collocated work over a period of time. The framework demonstrates how communicative action (as a factor) for coordination impacts BIM distributed and collocated work based on linguistic, representational, and structural forms. The team members recognize and assess the relevance of these forms to discuss, make decisions, and respond to requests from other actors (i.e., the relevance of forms for each stakeholder enables the deployment of the actions when participating in dialogs).

In this paper, the authors presented a nascent implementation of cognitive agent technology to increase BIM processes' efficiency to address coordination and challenges of understanding knowledge sharing among individual actors. It is envisioned that the cognitive agent will facilitate the enactment of conditions and practices to assist coordination in the BIM process. Introducing a cognitive agent design will respond to BIM technology challenges in collaborative practices when each team member (actor) views the collaborative design problem from a different 'lens' of reality. Embracing a broad cognitive agent implementation within the stakeholders' BIM process has the potential to improve design management processes, schedule efficiency, and constructability. The concept, named BIMbot, is a cognitive assistant, which future implementations will provide actionable information for coordination of activities, information, and advice on activities, engagement in teamwork by members executing a task, facilitation of fundamental actions for shared understandings as well as physical support and informed advice design.

Some aspects of technology implementation are worth emphasizing: Building an intelligent cognitive assistant is a challenging process. The technology design requires addressing dialogues to incorporate context. The context implies ambiguity and complexity. The use of natural language further increases such ambiguity and complexity – words are ambiguous in a different context for different users. Data availability is another significant challenge. There should be enough data to train the deep learning models to provide the information with less ambiguity. Retrieving information on domain-specific data makes it challenging to model due to the limited availability of information. There should be a combination of domain-specific and project-specific information for training and modelling to arrive at more useful information with less ambiguity for regular use in the BIM process and retrieve the most relevant information within any other question-and-answer dialogs. This approach used deep learning for neural machine translation to extract the query, and it used an information retrieval engine to best rank the relevance of the schema-based BIM software. Attempts to use rule-based models are easy to create, but they do not work well as they function purely on pattern matching. If there are no patterns found in response to a given query, then rule-based retrieval does not generate a response. Also, as pattern matching functions on rules (like first-order logic rules), it becomes a tedious task to write rules to cover all scenarios (Cahn, 2017).

The use of a cognitive assistant in the BIM process is a promising research area that would potentially reduce inefficiencies in the pre-design, design, and building processes for a construction project. It is expected that future developments will enable implementation for team members to interact with the cognitive agent to rapidly respond to queries and responses, thereby supporting collaboration and communication tasks and organizing work to achieve higher productivity and agility (Coronado et al., 2018) in construction projects.

In the future, we would like to venture into two interesting paths. The first is deploying the cognitive agent during a BIM meeting in a specific project to check the productivity and efficiency it brings during the meeting about a BIM process. The second path is to roll out the next version of our cognitive agent, which will be an upgrade from the current version. The next release will carry better features to solve project-specific problems, such as assisting in detecting clashes and an improved NLP engine with a syntactic and semantic analyzer.

REFERENCES

Addor, M. R. A. & Santos, E. T. BIM Design Coordination Room Infrastructure: Assessment of Communication Activities. Computing in Civil and Building Engineering, 2014, Orlando, FL. American Society of Civil Engineers, 1-8.

Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. International Conference on Learning Representations, ICLR 2015, 2015. San Diego, CA.

Beane, M. & Orlikowski, W. J. 2015. What Difference Does a Robot Make? The Material Enactment of Distributed Coordination. *Organization Science,* 26**,** 1553-1573.

Bechky, B. A. 2003. Sharing Meaning across Occupational Communities: The Transformation of Understanding on a Production Floor. *Organization Science,* 14**,** 312-330.

Bird, S., Klein, E. & Loper, E. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media.

Brownlee, J. 2017. *Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning*, Jason Brownlee.

Cahn, J. 2017. *CHATBOT: Architecture, design, & development.* University of Pennsylvania.

Carlile, P. R. 2002. A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. *Organization Science,* 13**,** 442-455.

Chen, M., Herrera, F. & Hwang, K. 2018. Cognitive Computing: Architecture, Technologies and Intelligent Applications. *IEEE Access,* 6**,** 19774-19783.

Cho, K., Van Merriënboer, B., Bahdanau, D. & Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *In:* Dekai Wu, M. C., Xavier Carreras, Eva Maria Vecchi, ed. Eighth Workshop on Syntax, Semantics, and Structure in Statistical Translation, September 2014, Doha, Qatar. Association for Computational Linguistics, 7.

Cmusphinx 2017. Open Source Speech Recognition Toolkit.

Coronado, M., Iglesias, C. A., Carrera, Á. & Mardomingo, A. 2018. A cognitive assistant for learning java featuring social dialogue. *International Journal of Human-Computer Studies,* 117**,** 55-67.

Crowston, K. 1997. A Coordination Theory Approach to Organizational Process Design. *Organization Science,* 8**,** 19.

Demirkan, H., Earley, S. & Harmon, R. R. 2017. Cognitive computing. *IT Professional,* 19**,** 16-20.

Dossick, C. S., Anderson, A., Azari, R., Iorio, J., Neff, G. & Taylor, J. E. 2014. Messy talk in virtual teams: Achieving knowledge synthesis through shared visualizations. *Journal of Management in Engineering,* 31**,** A4014003.

Eastman, C. M., Teicholz, P. M., Sacks, R. & Lee, G. 2018. BIM handbook: a guide to building information modeling for owners, managers, designers, engineers, and contractors. Third edition. Ed. Hoboken, New Jersey: Wiley.

Ghosh, T., Yates, J. & Orlikowski, W. J. 2004. Using Communication Norms for Coordination: Evidence from a Distributed Team.

Ghosh, T. Y., Joanne; and Orlikowski, Wanda 2004. Using Communication Norms for Coordination: Evidence from a Distributed Team. *International Conference on Information Systems (ICIS).* Washington, D.C.: ICIS.

Kaplan, S. & Orlikowski, W. J. 2013. Temporal Work in Strategy Making. *Organization Science,* 24**,** 965-995.

Karpathy, A. 2015. The Unreasonable Effectiveness of Recurrent Neural Networks. Available from: https://www.datasciencecentral.com/profiles/blogs/the-unreasonable-effectiveness-of-recurrent-neural-networks [Accessed March, 20 2019].

Le, N.-T. & Wartschinski, L. 2018. A Cognitive Assistant for improving human reasoning skills. *International Journal of Human-Computer Studies,* 117**,** 45-54.

Lecun, Y., Bengio, Y. & Hinton, G. 2015. Deep learning. *Nature,* 521**,** 436.

Li, Q. & Vasarhelyi, M. 2018. Developing a Cognitive Assistant for the Audit Plan Brainstorming Session. *The International Journal of Digital Accounting Research,* 18**,** 119-140.

Luong, T., Brevdo, E. & Zhao, R. 2017. Building Your Own Neural Machine Translation System in TensorFlow. *Google AI Blog* [Online]. [Accessed March, 19 2019].

Malone, T. W. & Crowston, K. 1990. What is coordination theory and how can it help design cooperative work systems? *Proceedings of the 1990 ACM conference on Computer-supported cooperative work.* Los Angeles, California, USA: Association for Computing Machinery.

Mehrbod, S., Staub-French, S., Mahyar, N. & Tory, M. 2019. Beyond The Clash: Investigating Bim-Based Building Design Coordination Issue Representation and Resolution. *Journal of Information Technology in Construction* 24**,** 24.

Merschbrock, C. & Munkvold, B. E. 2015. Effective digital collaboration in the construction industry–A case study of BIM deployment in a hospital construction project. *Computers in Industry,* 73**,** 1-7.

Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., Mcguinness, D. L., Morley, D., Pfeffer, A., Pollack, M. & Tambe, M. 2007. An intelligent personal assistant for task and time management. *AI Magazine,* 28**,** 47-47.

Okhuysen, G. A. & Bechky, B. A. 2009. Coordination in Organizations: An Integrative Perspective. *The Academy of Management Annals,* 3**,** 463-502.

Orlikowski, W. J. 1992. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science,* 3**,** 398-427

Orlikowski, W. J. 2007. Sociomaterial Practices: Exploring Technology at Work. *Organization Studies,* 28**,** 1435-1448.

Orlikowski, W. J. & Scott, S. V. 2008. Sociomateriality: Challenging the Separation of Technology, Work and Organization. *The Academy of Management Annals,* 2**,** 433-474.

Orlikowski, W. J. & Yates, J. 1994. Genre Repertoire: The Structuring of Communicative Practices in Organizations. *Administrative Science Quarterly,* 39**,** 541-574.

Pennington, J., Socher, R. & Manning, C. Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014 2014, Doha, Qatar. Association for Computational Linguistics, 1532-1543.

Preum, S. M., Munir, S., Ma, M., Yasar, M. S., Stone, D. J., Williams, R., Alemzadeh, H. & Stankovic, J. A. 2021. A Review of Cognitive Assistants for Healthcare: Trends, Prospects, and Future Directions. *ACM Comput. Surv.,* 53**,** Article 130.

Searle, J. R. 1969. *Speech acts: An essay in the philosophy of language* Cambridge, UK, Cambridge University Press.

Searle, J. R. 1995. *The construction of social reality,* New York, The Free Press.

Sennrich, R., Haddow, B. & Birch, A. Neural machine translation of rare words with subword units. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, Berlin, Germany. Association for Computational Linguistics, 1715–1725.

Siddike, M. A., Spohrer, J., Demirkan, H. & Kohda, Y. 2018. *People's Interactions with Cognitive Assistants for Enhanced Performances*.

Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014 2014. Cambridge, MA: MIT Press, 3104-3112.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q. & Macherey, K. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *Google Research* [Online], abs.

Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R. A new chatbot for customer service on social media. 2017 CHI Conference on Human Factors in Computing Systems, 2017 2017, Denver CO. ACM, 3506-3510.

Zachary C. Lipton, J. B., Charles Elkan 2015. A Critical Review of Recurrent Neural Networks for Sequence Learning. arXiv e-print.

Zasada, A. How Cognitive Processes Make Us Smarter. *In:* Di Francescomarino, C., Dijkman, R. & Zdun, U., eds. Business Process Management Workshops, 2019// 2019, Cham. Springer International Publishing, 45-55.