DOI: https://doi.org/10.20378/irb-46646
URN: urn:nbn:de:bvb:473-irb-466462

Intentional Forgetting in Artificial Intelligence Systems: Perspectives and Challenges

Ingo J. Timm[1], Steffen Staab[2], Michael Siebers[3], Claudia Schon[2], Ute Schmid[3], Kai Sauerwald[4], Lukas Reuter[1], Marco Ragni[5], Claudia Niederée[6], Heiko Maus[7], Gabriele Kern-Isberner[8], Christian Jilek[7], Paulina Friemann[5], Thomas Eiter[9], Andreas Dengel[7], Hannah Dames[5], Tanja Bock[8], Jan Ole Berndt[1], Christoph Beierle[4]*

[1] Trier University {itimm,reuter,berndt}@uni-trier.de
[2] University Koblenz-Landau {staab,schon}@uni-koblenz.de
[3] University of Bamberg {michael.siebers,ute.schmid}@uni-bamberg.de
[4] FernUniversität in Hagen
{kai.sauerwald,christoph.beierle}@fernuni-hagen.de
[5] Albert-Ludwigs-Universität Freiburg
{ragni,friemanp,damesh}@cs.uni-freiburg.de
[6] L3S Research Center Hannover niederee@l3s.de
[7] German Research Center for Artificial Intelligence (DFKI)
{christian.jilek,andreas.dengel,heiko.maus}@dfki.de
[8] TU Dortmund
tanja.bock@tu-dortmund.de,gabriele.kern-isberner@cs.uni-dortmund.de
[9] TU Wien eiter@kr.tuwien.at

Abstract. Current trends, like digital transformation and ubiquitous computing, yield in massive increase in available data and information. In artificial intelligence (AI) systems, capacity of knowledge bases is limited due to computational complexity of many inference algorithms. Consequently, continuously sampling information and unfiltered storing in knowledge bases does not seem to be a promising or even feasible strategy. In human evolution, learning and forgetting have evolved as advantageous strategies for coping with available information by adding new knowledge to and removing irrelevant information from the human memory. Learning has been adopted in AI systems in various algorithms and applications. Forgetting, however, especially intentional forgetting, has not been sufficiently considered, yet. Thus, the objective of this paper is to discuss intentional forgetting in the context of AI systems as a first step. Starting with the new priority research program on 'Intentional Forgetting' (DFG-SPP 1921), definitions and interpretations of intentional forgetting in AI systems from different perspectives (knowledge representation, cognition, ontologies, reasoning, machine learning, self-organization, and distributed AI) are presented and opportunities as well as challenges are derived.

Keywords: Artificial Intelligence Systems · Capacity and Efficiency of Knowledge-Based Systems · (Intentional) Forgetting.

*Author names appear in alphabetical order

1 Introduction

Today's enterprises are dealing with massively increasing digitally available data and information. Current technological trends, e.g., Big Data, focus on aggregation, association, and correlation of data as a strategy to handle information overload in decision processes. From a psychological perspective, humans are coping with information overload by selective *forgetting* of knowledge. Forgetting can be defined as non-availability of a previously known certain piece of information in a specific situation [29]. It is an adaptive function to delete, override, suppress, or sort out outdated information [4]. Thus, forgetting is a promising concept of coping with information overload in organizational contexts.

The need for forgetting has already been recognized in computer science [17]. In logics, context-free forgetting operators have been proposed, e.g., [6,30]. While logical forgetting explicitly modifies the knowledge base (KB), various machine learning approaches implicitly forget details by abstracting from their input data. In contrast to logical forgetting, machine learning can be used to reduce complexity by aggregating knowledge instead of changing the size of a KB. As a third approach, distributed AI (DAI) focuses on reducing complexity by distributing knowledge across agents [21]. These agents 'forget' at the individual level while the overall system 'remembers' through their interaction.

For humans, forgetting is *also an intentional mechanism* to support decision-making by focusing on relevant knowledge [4,22]. Consequently, the questions arise when and how humans can intentionally forget and when and how intelligent systems should execute forgetting functions. The new priority research program on "Intentional Forgetting in Organizations" (DFG-SPP 1921) has been initiated to elaborate an interdisciplinary paradigm. Within the program, researchers from computer science and psychology are interdisciplinarily collaborating on different aspects of intentional forgetting in eight tandem projects.[1]

With a strong focus (five projects) on AI systems, multiple perspectives are researched ranging from knowledge representation, cognition, ontologies, reasoning, machine learning, self-organization, and DAI. In this paper we bring together these perspectives as a first building block for establishing a common understanding of intentional forgetting in AI. Contributions of this paper are the identification of AI research fields and their challenges.

2 Knowledge Representation and Cognition: FADE

The goal of FADE (Forgetting through Activation, reDuction and Elimination) is to support the effortful preselection and aggregation of information in information flows, leading to a reduction of the user's workload, by integrating methods from cognitive and computer science: Knowledge structures in organizations and mathematical and psychological modeling approaches of human memory structures in cognitive architectures are analyzed. Functions for priorization and forgetting that may help to compress and reduce the increasing

[1] `http://www.spp1921.de/projekte/index.html.de`

amount of data are designed. Furthermore, a cognitive computational system for forgetting is developed that offers the opportunity to determine and adapt system model parameters systematically and makes them transparent for every single knowledge structure. This model for forgetting is evaluated for its fit to a lean workflow and readjusted in the context of the ITMC of the TU Dortmund.

While forgetting is often attributed negatively in everyday life, forgetting can offer an effective and beneficial reduction process to allow humans to focus on information of higher relevance. Features of the cognitive forgetting process which are crucial to the FADE project work are that information never gets lost but instead has a level of activation [1], and that the relevance of information depends on its connection to other information and its past usage. Moreover, information characteristics require different forms of forgetting; in particular, insights from knowledge representation and reasoning can help to further refine declarative knowledge, and differentiate between assertional knowledge and conceptual knowledge. Finally, it can be expected that cognitive adequacy of forgetting approaches will improve the human-computer interaction significantly.

The project FADE focusses on formal methods that are apt to model the epistemic and subjective aspects of forgetting [3,13]. Here, the wide variety of formalisms of nonmonotonic reasoning and belief revision are extremely helpful [2]. The challenge is to adapt these approaches to model human-like forgetting, and to make them usable in the context of organizations. As a further milestone, these adapted formal methods are integrated into cognitive architectures providing a formal-cognitive frame for forgetting operations [23,24].

3 Ontologies and Reasoning: EVOWIPE

New products are often developed by modifying the model of an already existing product. Assuming that large parts of the product model are represented in a KB, the EVOWIPE project supports this reuse of existing product models by providing methods to intentionally forget aspects from a KB that are not applicable to the new product [14]. E.g., the major part of the product model of the VW e-Golf (with electric motor) is based on the concept of the VW Golf with combustion engine. However, (i) changes, (ii) additions and (iii) forgetting elements of the original product model are necesarry, e.g. (i) connecting the engine, (ii) adding a temperature control system for the batteries, and (iii) forgetting the fuel tank, fuel line and exhaust gas treatment. EVOWIPE aims at developing methods to support the product developer in the process of forgetting aspects from product models represented in KBs by developing the following operators for intentional forgetting: *Forgetting of inferred knowledge, restoring forgotten elements, temporary forgetting, representation of place markers in forgetting, cascading forgetting.*

These operators bear similarities to deletion operators known in knowledge representation (cf. Section 2). Indeed, we represent knowledge about product models by transforming existing product model data structures into an OWL-based representation and build on existing research that accesses such KBs using

SPARQL update queries. These queries allow not only for deleting knowledge but also for inserting new knowledge. Therefore, the interplay of deletion and insertion is investigated in the project as well [25]. To accomplish cascading forgetting, dependencies occurring in the KB have to be specified. They can be added as metaproperties into the KB [10]. These dependencies can be added manually, however the project partners are currently working on methods to automatically extract dependencies from the product model. Dependency-guided semantics for SPARQL update queries use these dependencies to accomplish the desired cascading behavior described above [15]. By developing these operators, the EVOWIPE project extends the product development process to include stringent methods for intentional forgetting, ensuring that the complexity inherent in the product model, the product development process and the forgetting process itself can be mastered by the product developer.

4 Machine Learning: Dare2Del

Dare2Del is a system designed as context-aware cognitive companion [9,26] to support forgetting of digital objects. The companion will help users to delete or archive digital objects which are classified as irrelevant and it will support users to focus on a current task by fading-out or hiding digital information which is irrelevant in a given task context. In collaboration with psychology, it is investigated for which persons and in which situations information hiding can improve task performance and how explanations can establish trust of users in system decisions. The companion is based on inductive logic programming (ILP) [18] – a white-box machine learning approach based on Prolog. ILP allows learning from small sets of training data, a natural combination of reasoning and learning, and the incorporation of background knowledge. ILP has been shown to be able to provide human-understandable classifiers [19].

For Dare2Del to be a cognitive companion, it should be able to explain system decisions to users and be adaptive. Therefore, we currently design an incremental variant of ILP to allow for interactive learning [8]. Dare2Del will take into account explanations given by the user. E.g., if a user decides that an object should not be deleted, he or she can select one or more predicates (presented in natural language) which hold for the object and which are the reason why it should not be deleted. Subsequently, Dare2Del has to adapt its model. As application scenarios for Dare2Del we consider administration as well as connected industry. In the context of administration, users will be supported to delete irrelevant files and Dare2Del will help to focus attention by hiding irrelevant columns in tables. In the context of connected industry, quality engineers are supported in identifying irrelevant measurements and irrelevant data for deletion. Alternatively, measurements and data can be hidden in the context of a given control task. We believe that Dare2Del can be a helpful companion to relieve humans from the cognitive burden of complex decision making which is often involved when we have to decide whether some digital object will be relevant in the future or not.

5 Self-Organization: Managed Forgetting

We investigate intentional forgetting in grass-roots (i.e. decentralized and self-organizing) organizational memory, where knowledge acquisition is incorporated into daily activities of knowledge workers. In line with this, we have introduced *Managed Forgetting* (MF) [20] - an evidence-based form of intentional forgetting, where no explicated will is required: what to forget and what to focus on is learned in a self-organizing and decentralized way based on observed evidences.

We consider two forms of MF: *memory buoyancy* empowering forgetful information access and *context-based inhibition* easing context switches. We apply MF in the *Semantic Desktop*, which semantically links information items in a machine understandable way based on a *Personal Information Model* (PIMO) [11]. Shared parts of individual PIMOs form a basis for an Organizational Memory.

As a key concept for this form of MF we have presented *Memory Buoyancy* (MB) [20], which represents an information item's current value for the user. It follows the metaphor of less relevant items "sinking away" from the user, while important ones are pushed closer. MB value computation has been investigated for different types of resources [5,28] and is based on a variety of evidences (e.g. user activities), activation propagation as well as on heuristics. MB values provide the basis for **forgetful access methods** such as hiding or condensation [11], adaptive synchronization and deletion, and forgetful search.

Most knowledge workers experience frequent context switches due to multitasking. Other than the gradual changes of MB in the first form of MF, in the case of context switches, changes are far more abrupt. We, therefore, believe that approaches based on the *concept of inhibition* [16], which temporarily hide resources of other contexts could be employed here, e.g. in a kind of self-tidying and self-(re)organizing context spaces [12]. Our current research focuses on combining both forms of MF.

6 Distributed Artificial Intelligence: AdaptPRO

In DAI, (intelligent) agents encapsulate knowledge which is deeply connected to domain, tasks and action [21]. They are intended to perceive their environment, react to changes, and act autonomously by (social) deliberation. Forgetting is implicitly a subject of research, e.g., Belief Revision (cf. section 2) or possible-worlds semantics [31]. By contrast, the team perspective of forgetting, i.e., change of knowledge distribution, roles, and processes have not been analyzed yet.

In AdaptPRO, we focus on these aspects by adopting intentional forgetting in teams from psychology. We define *intentional forgetting* as the reorganization of knowledge in teams. The organization of human team knowledge is known as team cognition (TC). TC describes the structure in which knowledge is mentally represented, distributed, and anticipated by members to execute actions [7]. The concept of TC can be used to model knowledge distribution in agent systems as well. In terms of knowledge distributions, organization of roles and processes are implemented by allocating, sharing or dividing knowledge. If certain team members are specialized on particular areas, other agents can ignore information related to this area [27]. Especially, when cooperating, it is important for agents to share their knowledge about task- and team-relevant information. Particularly in case of disturbances, redundant knowledge and task competences enable robust teamwork. To strike a balance between sharing and dividing knowledge, i.e., efficient and robust teamwork, AdaptPRO applies an interdisciplinary approach of modeling, analyzing and adapting knowledge structures in teams and measure their implications on individual and team perspective.

7 Challenges and Future Work

We have presented perspectives on intentional forgetting in AI systems. Their key opportunities can be summarized as follows: (a) Establishing guidelines that help to implement human-like forgetting for organizations by bridging Cognition and Organizations with formal AI methods. (b) Mastering information overload by (temporary) forgetting and restoring of knowledge with respect to inferred and cascading knowledge structures. (c) Supporting decision-making of humans by forgetting digital objects with comprehensive knowledge management and machine learning. (d) Assisting organizational knowledge management with intentional forgetting by self-organization and self-tidying. (e) Adapting processes and roles in organizations by reorganization of knowledge distribution. In order to tap into these opportunities, the following challenges must be overcome: (1) Merge concepts of (intentional) forgetting in AI in a common terminology. (2) Formalize kinds of knowledge and forgetting to make prerequisites and aims of forgetting operations transparent and study their formal properties. (3) Investigate whether different forms of knowledge require different techniques of forgetting. (4) Accomplish efficient remembering of knowledge. (5) Develop temporarily forgetting information from a KB. (6) Develop of an incremental probabilistic approach to inductive logic programming which allows interactive learning by mutual explanations. (7) Generate helpful explanations in form of verbal justifications and by providing examples or counterexamples. (8) Develop correct interpretation on user activities, work environment, and information to initiate appropriate forgetting measures. (9) Characterize knowledge in teams and DAI-Systems and develop formal operators for reallocating, extending, and forgetting information.

These challenges foster an important basis for AI research in the next years. Furthermore, intentional forgetting has the potential to evolve to a mandatory function of next generation AI systems, which become capable of coping with our days' complexity and data availability.

Acknowledgments. The authors are indebted to the DFG for funding this research: Dare2Del (SCHM1239/10-1), EVOWIPE (STA572/15-1), FADE (BE 1700/9-1, KE1413/10-1, RA1934/5-1), Managed Forgetting (DE420/19-1, NI1760 1-1), and AdaptPro (TI548/5-1). We would also like to thank our project partners for their fruitful discussion: C. Antoni, T. Ellwart, M. Feuerbach, C. Frings, K. Göbel, P. Kügler, C. Niessen, Y. Runge, T. Tempel, A. Ulfert, S. Wartzack.

References

1. Anderson, J.R.: How can the human mind occur in the physical universe? Oxford University Press, New York (2007)
2. Beierle, C., Kern-Isberner, G.: Semantical investigations into nonmonotonic and probabilistic logics. Annals of Mathematics and Artificial Intelligence **65**(2), 123–158 (2012)
3. Beierle, C., Eichhorn, C., Kern-Isberner, G.: Skeptical inference based on c-representations and its characterization as a constraint satisfaction problem. In: Foundations of Information and Knowledge Systems - 9th International Symposium, FoIKS 2016, Linz, Austria, March 7-11, 2016. Proceedings. Lecture Notes in Computer Science, vol. 9616, pp. 65–82. Springer (2016). https://doi.org/10.1007/978-3-319-30024-5_4
4. Bjork, E.L., Anderson, M.C.: Varieties of goal-directed forgetting. In: Golding, J.M., MacLeod, C. (eds.) Intentional Forgetting: Interdisciplinary Approaches, pp. 103–137. Lawrence Erlbaum: Mahwah, NJ (1998)
5. Ceroni, A., Solachidis, V., Niederée, C., Papadopoulou, O., Kanhabua, N., Mezaris, V.: To keep or not to keep: An expectation-oriented photo selection method for personal photo collections. In: Proc. of the 5th ACM Int'l Conf. on Multimedia Retrieval, Shanghai, China, June 23-26, 2015. pp. 187–194 (2015)
6. Delgrande, J.P.: A knowledge level account of forgetting. J. Artif. Intell. Res. **60**, 1165–1213 (2017)
7. Ellwart, T., Antoni, C.H.: Shared and distributed team cognition and information overload. evidence and approaches for team adaptation. In: Marques, R.P.F., Batista, J.C.L. (eds.) Information and Communication Overload in the Digital Age, pp. 223–245. IGI Global, Hershey (2017)
8. Fails, J.A., Olsen Jr, D.R.: Interactive machine learning. In: Proceedings of the 8th International Conference on Intelligent User Interfaces. pp. 39–45. ACM (2003)
9. Forbus, K.D., Hinrichs, T.R.: Companion cognitive systems: A step toward human-level AI. AI magazine **27**(2), 83 (2006)
10. Guarino, N., Welty, C.A.: An overview of ontoclean. In: Handbook on Ontologies, pp. 201–220. International Handbooks on Information Systems, Springer (2009)
11. Jilek, C., Maus, H., Schwarz, S., Dengel, A.: Diary generation from personal information models to support contextual remembering and reminiscence. In: 2015 IEEE Int'l Conf. on Multimedia & Expo Workshops, ICMEW 2015. pp. 1–6 (2015)
12. Jilek, C., Schröder, M., Schwarz, S., Maus, H., Dengel, A.: Context spaces as the cornerstone of a near-transparent & self-reorganizing semantic desktop. In: The Semantic Web: ESWC 2018 Satellite Events. Springer (2018)
13. Kern-Isberner, G., Bock, T., Sauerwald, K., Beierle, C.: Iterated contraction of propositions and conditionals under the principle of conditional preservation. In: Benzmüller, C., Lisetti, C.L., Theobald, M. (eds.) GCAI 2017, 3rd Global Conference on Artificial Intelligence, Miami, FL, USA, 18-22 October 2017. Proceedings. EPiC Series in Computing, vol. 50, pp. 78–92. EasyChair (2017), `http://www.easychair.org/publications/paper/DTmX`
14. Kestel, P., Luft, T., Schon, C., Kügler, P., Bayer, T., Schleich, B., Staab, S., Wartzack, S.: Konzept zur zielgerichteten, ontologiebasierten Wiederverwendung von Produktmodellen. In: Krause, D., Paetzold, K., S. Wartzack, S. (eds.) Design for X. Beiträge zum 28. DfX-Symposium. pp. 241–252. TuTech Verlag, Hamburg (2017)

15. Kügler, P., Kestel, P., Schon, C., Marian, M., Schleich, B., Staab, S., Wartzack, S.: Ontology-based approach for the use of intentional forgetting in product development. In: DESIGN Conference Dubrovnik (2018)
16. Levy, B.J., Anderson, M.C.: Inhibitory processes and the control of memory retrieval. Trends in cognitive sciences **6**(7), 299–305 (2002)
17. Markovitch, S., Scott, P.D.: Information filtering: Selection mechanisms in learning systems. Machine Learning **10**(2), 113–151 (Feb 1993)
18. Muggleton, S., De Raedt, L.: Inductive logic programming: Theory and methods. The Journal of Logic Programming **19**, 629–679 (1994)
19. Muggleton, S.H., Schmid, U., Zeller, C., Tamaddoni-Nezhad, A., Besold, T.: Ultra-strong machine learning-comprehensibility of programs learned with ilp. Machine Learning (2018)
20. Niederée, C., Kanhabua, N., Gallo, F., Logie, R.H.: Forgetful digital memory: Towards brain-inspired long-term data and information management. SIGMOD Record **44**(2), 41–46 (2015)
21. O'Hare, G.M.P., Jennings, N.R. (eds.): Foundations of Distributed Artificial Intelligence. John Wiley & Sons, Inc., New York (1996)
22. Payne, B.K., Corrigan, E.: Emotional constraints on intentional forgetting. Journal of Experimental Social Psychology **43**(5), 780–786 (2007)
23. Ragni, M., Sauerwald, K., Bock, T., Kern-Isberner, G., Friemann, P., Beierle, C.: Towards a formal foundation of cognitive architectures. In: Proceedings of the 40th Annual Meeting of the Cognitive Science Society, CogSci 2018, Madison, US, 25-28 July 2018 (2018), (to appear)
24. Sauerwald, K., Ragni, M., Bock, T., Kern-Isberner, G., Beierle, C.: On a formalization of cognitive architectures. In: Proceedings of the 14th Biannual Conference of the German Cognitive Science Society. Darmstadt (2018), (to appear)
25. Schon, C., Staab, S.: Towards SPARQL instance-level update in the presence of OWL-DL tboxes. In: JOWO. CEUR Workshop Proceedings, vol. 2050. CEUR-WS.org (2017)
26. Siebers, M., Göbel, K., Niessen, C., Schmid, U.: Requirements for a companion system to support identifying irrelevancy. In: International Conference on Companion Technology, ICCT 2017, Ulm, Germany, September 11-13, 2017. pp. 1–2. IEEE (2017)
27. Timm, I.J., Berndt, J.O., Reuter, L., Ellwart, T., Antoni, C., Ulfert, A.S.: Towards multiagent-based simulation of knowledge management in teams. In: Leyer, M., Richter, A., Vodanovich, S. (eds.) Flexible knowledge practices and the Digital Workplace (FKPDW). Workshop within the 9th Conference on Professional Knowledge Management. pp. 25–40. KIT: Karlsruhe (2017)
28. Tran, T., Schwarz, S., Niederée, C., Maus, H., Kanhabua, N.: The forgotten needle in my collections: Task-aware ranking of documents in semantic information space. In: CHIIR-16. ACM Press (2016)
29. Tulving, E.: Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. American Scientist **62**(1), 74–82 (1974)
30. Wang, Z., Wang, K., Topor, R., Pan, J.Z.: Forgetting for knowledge bases in dl-lite. Annals of Mathematics and Artificial Intelligence **58**(1), 117–151 (Feb 2010)
31. Werner, E.: Logical foundations of distributed artificial intelligence. pp. 57–117. John Wiley & Sons, Inc., New York (1996)