OVAL: Open Virtual Assistant Lab

Monica S. Lam
Computer Science Department
Stanford University

1 Executive Summary

The Open Virtual Assistant Lab is formed to advance virtual assistant technology and to bring companies together to create an industrial ecosystem that protects consumer privacy and promotes open competition.

The Problem. Linguistic computer interfaces are revolutionizing how humans connect with machines in the mobile and ubiquitous computing era. Open-world collaboration is necessary to teach computers how to assist users in every domain and in every human language. We wish to accelerate the development of this technology and make it available to every company and nonprofit organization.

Linguistic interface is most prominently used in virtual assistants today, which provides the compelling value proposition of giving users a unified, personalized linguistic interface to all their IoT and web accounts. These assistants are in the position to collect a massive amount of user data and to control linguistic access to the web. Due to the high cost in developing natural language technology, and due to the network effect, virtual assistants will likely become a platform oligopoly unless there is a successful open alternative.

The Team. We have assembled a diverse team of experts at Stanford in computer science and law. To solve this important real-world problem, we have assembled a world-class team of experts in computer science and law: Michael Bernstein (Crowdsourcing) Dan Boneh (Security), Jen King (Consumer Privacy Law), Monica Lam (Programming Systems, AI), James Landay (Human-computer interaction), Fei-fei Li (AI), Chris Manning (AI, NLP), David Mazières (Blockchains), Chris Re (Database, AI).

The Research. Through collaboration with industry, our goal is to develop fundamental technology, create and evangelize open standards, develop product-quality infrastructures, and to assist in product development on our infrastructure.

1. Conduct research on linguistic and multimodal interface technologies.
2. Create a WORA (Write Once and Run on Any virtual assistant) Skill platform. This will eliminate duplication of effort and level the playing field for virtual assistants.
3. Create the best and open neural model for language interfaces so companies can own and deploy their lingistic interfaces.
4. Create private, secure assistants capable of conducting confidential transactions in medical and financial domains.

5. Create standard protocols to support interoperability across virtual assistants to prevent fragmented oligopolies and accelerate the growth of the linguistic web.

6. Create blockchain-based technology that lets users control third-party sharing of personal data conveniently between institutions. This is particularly useful for sharing of health, finance, and education data.

In three years' time, we expect to establish Open Virtual Assistants as a commercial alternative to existing systems and to create a non-profit industry consortium open to all companies. This consortium should be of interest to consumer businesses (retailers, hardware, autos, hotels, health and finance industries), as well as all professions whose workflow can be improved with virtual assistants.

Engagement with Industry. OVAL is supported by unrestricted funding from partner companies that each contribute $500K per year plus additional support from the National Science Foundation. Companies joining the OVAL affiliate program are invited to contribute to the open-source research of virtual assistant technology and the creation of open standards. They are welcome to participate in annual retreats, workshops, and seminars. Members will have the opportunity to get first-hand information on the work in progress and influence research directions. Such interactions often precipitate joint projects between Stanford researchers and one or more companies.

2 Motivation

The virtual assistant will transform our digital experience by giving us a fully personalized and integrated linguistic interface to our digital assets, which are currently siloed in different services. Furthermore, as it collects and analyzes detailed information across all users, it will learn an accurate model of human behavior. It will predict and intervene with our behavior, such as reminding us to take our medication. Similarly, by gathering details on business decisions and outcomes, assistants will monitor and optimize business logic across professions.

It is likely that a platform monopoly or duopoly will emerge for the virtual assistant. Monopolies hurt consumers as they stifle competition and innovation. A virtual assistant monopoly is particularly worrisome because it controls consumers' access to the digital world, and sees the private data of billions of people across all different services.

A proprietary linguistic web. We use the virtual assistant to access the linguistic web, just like how we use the browser to access the graphical web. However, while the graphical web is non-proprietary, we are witnessing the creation of proprietary linguistic webs. It is hard to understand one natural language, let alone the many languages needed for the international market. The state-of-the-art neural network approach requires a large volume of annotated human sentences; hence, Amazon has 10,000 employees devoted to Alexa [9]. Moreover, popular assistants attract device manufacturers and service providers, which, in turn, attract more users. Thus, a significant barrier to entry, due to the network effect and high development cost, will likely lead to a virtual assistant oligopoly.

By intermediating between consumers and the web, virtual assistants have the power to channel users to their own or promoted products. They may even charge a fee to commercial transactions conducted on their platform, similar to App Stores charging mobile apps 30% of their revenues. More importantly, by getting access to users' accounts, virtual assistants are privy to valuable business intelligence data. For example, users' thermostat accounts contain valuable energy usage data. A virtual assistant monopoly will have an unfair advantage in all consumer businesses.

Monopoly platforms threaten privacy. There is a growing awareness in both the United States and the European Union that large platforms pose severe threats to individual privacy. The General Data Protection Regulation by the E.U. represents the first systematic attempt to address the existing dominance large platforms have over personal information, allowing users to transfer data from one platform to another. Companies owning billions of people's personal information have tremendous power; they can share users' data with other parties, influence users' opinion in important decisions such as presidential elections, and intervene with users' behavior. For example, the social networking market is dominated globally by Facebook, which has no meaningful competition today in most countries. A monopoly assistant platform will have access to data in all our different accounts; it will have more knowledge than Amazon, Facebook, and Google combined.

3 An Open Virtual Assistant Manifesto

It is essential that we lower the barrier to entry to virtual assistants, increase innovation and open competition, and give consumers a choice. We can summarize our vision with the following manifesto:

1. Democratize AI for linguistic user interfaces. We should have open, collaborative research to put the best linguistic technology in the hands of all businesses.

2. An open, non-proprietary linguistic web. All skills, or linguistic user interfaces, should be made available to any virtual assistant.

3. Sharing with individual data ownership. Consumers should have a choice in virtual assistant services and the ability to control how we share our data.

4 Technical Foundation

In our four years of research *prior* to receiving the NSF grant, we have greatly expanded the capability of existing virtual assistants and reduced the cost in creating interfaces. All the work has been demonstrated in a fully working research prototype, called Almond [3], which has been publicly released on GitHub [1] and on the web [2]. An overview of the Almond architecture is shown in Figure 4.

Natural language programming. Today's existing virtual assistants can handle mainly simple, hardcoded commands, such as turning on the TV, or getting information from a single source such as "Ask Bing to search for pizza". A true assistant will help the user with applying information from one domain to another, such as taking a restaurant recommendation and checking for availability in OpenTable, or sharing information with friends, or making decisions. For example, we want to monitor the price of a given stock and buy some shares when the price drops to a specified level. Users can get such assistance with our technology by using natural language to *describe* tasks crossing over domains. The innovation is to train a neural network to translate natural language into ThingTalk, a domain-specific language we invented to facilitate natural language translation. The skill representation in the Thingpedia repository is a superset of that in commercial assistants because it includes the full API signature rather than just intents and input parameters. Our representation uniquely supports compositionality, which is critical for natural language programming.

The Genie linguistic interface generator. Neural networks are notorious for needing lots of training data. How do we train our neural network for new capabilities, such as natural language programming, when such commands have never been written? We have created a tool called Genie

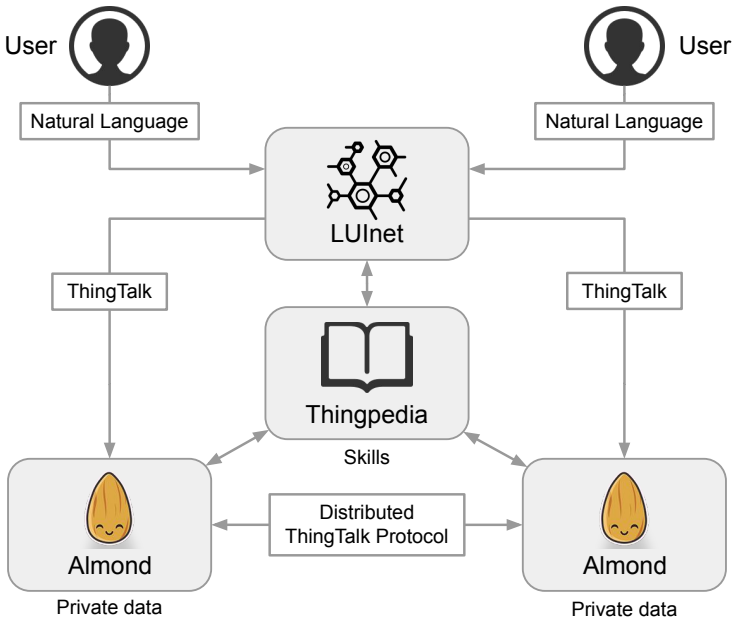

Figure 1: Each user has their private instance of Almond, storing their data. These private instances make use of the shared LUInet and Thingpedia, which are grown collaboratively and contain the information to translate the natural language to executable ThingTalk code. Different instances of Almond can share data through a distributed protocol.

that lets domain experts create a starter linguistic interface by engineering a training set. Genie requires relatively little manual effort and no in-house machine-learning expertise [4]. We have trained a neural network called LUInet that understands the composition of commands over more than 100 skills and 250 functions. The LUInet model and training data are publicly available, and are continuously updated as new skills are added to Thingpedia.

Sharing with privacy. We have shown that virtual assistants can be our trusted agent to help us share personal data with full privacy, all without disclosing any information to a third party [5]. In addition, we can also control companies in how they share their data with human-understanding agreements kept in auditable and indelible records. We have created a distributed protocol for sharing ThingTalk commands to enable interoperability among virtual assistants. This supports sharing without centralization.

5 Research Projects in the Lab

Open-source technology has shown to be effective as an alternative to monopolies. Unix, first released in 1971, and subsequently BSD and Linux offered an alternative to Windows and led to the server and mobile OSes used widely today. The NCSA Mosaic browser, first released in 1993, and subsequently Netscape and Firefox offered a widely adopted open-source alternative to the Internet Explorer. Recently, the success of AlexNet, VGG, Inception, ResNet for object recognition demonstrates the importance of open research in machine learning. In the following, we describe some of our open-source research projects in progress.

Conversational Agent Research. Instead of expecting the user issue commands after commands, the future virtual assistant will engage users in a user-friendly dialog. Multi-turn dialogs are challenging; we plan to model partial dialogs with a formal representation that is fed back as context to the neural model. We will also investigate how agents can speak in a way that people find natural and enjoyable, building upon previous work in the Engagement Learning Interaction Agent (ELIA) [8]. In contrast to active learning, we propose *engagement learning*, where we trade off what the AI needs, the *knowledge value* of the label to the AI model, against what people are interested in, the *engagement value* of the label. The research will include open-vocabulary interactions, and will involve interacting with people directly.

We will study how to mix the voice and graphical interfaces to get the best of both worlds. The graphical interface makes it easy to view images, to monitor multiple queries simultaneously, to re-run complex commands, and to adjust settings using graphical widgets [7]. By mixing these modes, we can conveniently switch context, especially on mobile phones, where there is limited screen real estate.

Today, all our data are silo'ed in different services and users are given limited access. Thanks to GDPR, companies are required to release all their data to the owners. One of our research goals is to build a virtual assistant that lets users query all their personal data from all the web services using natural language. They can aggregate information such as asking for the total expenditure across all their credit cards, or share any data with whomever they wish. For example, a Facebook user can use a virtual assistant to share his data with non-Facebook friends; or a Facebook user can download his data, delete his account, and still access it easily in natural language.

We will explore how to present privacy and consent notices from a human-centered perspective that respects individuals' privacy preferences and is compliant with privacy laws. We will perform user studies on the conversational aspects of managing privacy in daily life, identify and analyze real-life discussions of privacy preferences and concerns, and implement our findings in the Almond assistant. The result of this work will include a taxonomy of privacy concerns, heuristics, and negotiative language mapped to legal concepts to incorporate into natural-language based platforms, as well as data regarding the public's general and context-specific privacy and data use concerns with AI-powered virtual assistants.

Building the Best Neural Model with Real-Life Deployment. We estimate that it will take five years to create a LUInet that is proficient in understanding verbal instruction of all digital tasks. This task requires us to develop industry-strength tools that can help companies create interfaces for their products, and collect real-life utterances to jointly train LUInet. Research has shown that training for multiple domains all at once can improve the accuracy of individual domains [6, 11]. By accumulating contributions from experts in different domains, we can create an open LUInet that is superior to any proprietary model developed by one company.

An Open Skill Service. Today, commercial skill platforms for Alexa and Google Assistant are open to skill developers, but they are proprietary, meaning that Amazon and Google have full control over the access to these skills. We plan to create a "write once, run anywhere" WORA Skill Service so developers only need to enter their information once, and we will make it automatically available to Alexa, Google Assistant, and any other assistants. This can be achieved easily because our Thingpedia representation supersedes those in commercial platforms, and it uniquely supports compositionality. Not only would developers find this WORA Skill Service convenient, doing so can level the playing field in virtual assistants, reduce their dependence on dominant platforms, and open up competition among virtual assistant vendors. More importantly, unlike Alexa and Google Assistant, this platform makes the AI (the neural model and linguistic interface) available to the

skill provider so they can incorporate that into their own website, app, or phone service, as shown in Figure 5.

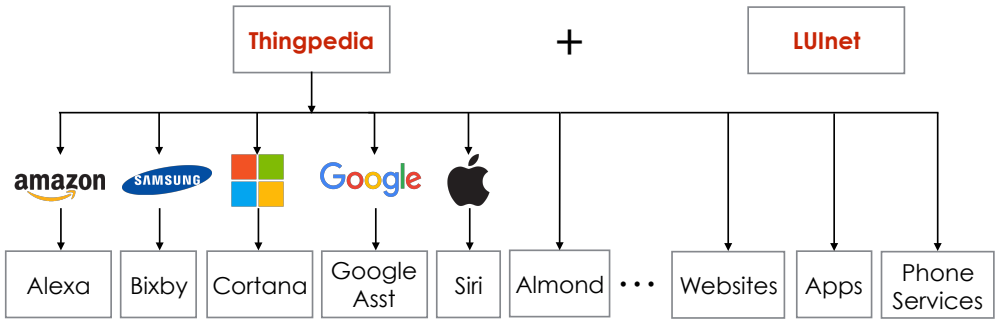

Figure 2: Skills in Thingpedia are WORA (Write Once and Run on Any assistant) and the neural model for the linguistic interface LUInet is available to be embedded in the skill provider's websites, apps, and phone services.

Privacy-Honoring Assistants. While virtual assistants are useful for financial, health and education services, such industries cannot deploy their services on commercial virtual assistants that do not protect user privacy. We plan to develop a production-quality Almond virtual assistant specifically for privacy-sensitive use cases. Our first use case, real-time monitoring of patients' blood pressures, will be studied under the guidance of Dr. Alan Yeung, a cardiologist from Stanford Medical School. In this use case, Dr. Yeung will prescribe personalized blood-pressure monitoring schedules for each of his patients; the Almond assistant will automatically remind his patients to submit their results and notify him of all readings exceeding the individually designed thresholds. The assistant allows Dr. Yeung to provide better health care to those in need, while eliminating unnecessary doctor visits and increasing the number of patients he can serve. We expect that many professionals can benefit similarly from virtual assistants that can be entrusted with access to corporate-confidential APIs.

Secure Third-Party Sharing Across Institutions With Blockchain Technology. As consumers' personal data are held in various institutions, it is important that they can authorize sharing of data, and to control and monitor how their data are shared with third parties. This is especially important for medical, financial, and education data which are highly regulated by law. The sharing contract language must be rich enough to reflect the kind of control individuals wish to impose. For example, patients may allow their records to be shared with CDC provided their personal identification information is removed. Instead of legalese, these contracts must be in natural language that can be understood. Contracts need to be negotiable in natural language dialogs, and must be enforceable directly and be revocable. Data must be shared directly, and not via a centralized facility, with all transactions recorded in an indelible and auditable trail.

Our first prototype is MedXchange, a platform developed in collaboration of Dr. Lei Xing, a radiation oncologist at Stanford, that lets users control how institutions share their medical images with third parties. For auditability, we keep a hash of the contracts and sharing transactions on an efficient blockchain using federated Byzantine agreements[10]. The successful deployment of such a system can increase the availability of health data along with patients' consent for medical research.

6 Research Team

Stanford has a track record in disrupting platform monopolies. In 2008, Stanford was awarded a $10M NSF Grant "Programmable Open Mobile Internet (POMI) 2020" with the goal to disrupt the Cisco monopoly and promote innovation in the networking industry. The project started out as a Stanford research project, it evolved to become a collaborative effort between Stanford and major players such as Deutsche Telekom, Google, and NEC, and it is now supported by three open industry consortia consisting of over 150 companies. The result is an open programmable standard in communications called Software Defined Networking (SDN); it opened up the competition, invigorated investment in new communications startups, and has been projected to become a $100B industry by 2025.

Prof. Monica Lam, the Principal Investigator in the Open Virtual Assistant Lab, was a co-PI in the POMI research project. We plan to follow the same 3-step approach as SDN. First, we have obtained $3M funding from National Science Foundation in April 2019. Second, we are starting the OVAL affiliate program to bring together companies to advance the state of the art, to define standards for open and interoperable virtual assistants and skills, and to deploy this technology. Third, we plan to create an industry consortium open to all companies in about three years.

Investigator	**Field**	**Awards**
Michael Bernstein	Crowdsourcing	NSF Career
Dan Boneh	Security	NAE, ACM Fellow, Godel prize
Jen King	Director of Consumer Privacy (Law School)	
Monica Lam (PI)	Programming Systems, AI	NAE, ACM Fellow
James Landay	Human-Computer Interaction	ACM Fellow, CHI Academy
Fei-fei Li	AI	ACM Fellow
Chris Manning	Natural language Processing, AI	ACM, AAAI, ACL Fellow
David Mazières	Blockchain	NSF Career, Stellar Co-founder
Chris Re	Knowledge Bases, AI	MacArthur Fellow

7 Conclusion

Natural language is the new user interface for the era of mobile and ubiquitous computing. The current trends point to a likely emergence of an oligopoly platform that will intermediate linguistic interactions between consumers and businesses. This will pose a huge threat to consumer privacy and open competition. We need to make natural language technology open and available to all industry, to keep the linguistic web open, and to create a privacy-minded virtual assistant that finance and health industries can trust. We have built a world-class team, laid a technical foundation for the open assistant, and developed a concrete roadmap to democratize AI and protect privacy.

References

[1] Giovanni Campagna, Michael Fischer, Johnny Hsu, Monica S. Lam, Alison Lin, Elvis Yu-Jing Lin, Wesley Liu, Mehrad Moradshahi, Rakesh Ramesh, Silei Xu, Jackie Yang, and Richard Yang. GitHub - Stanford Open Virtual Assistant Lab. `https://github.com/stanford-oval`, 2019.

[2] Giovanni Campagna, Michael Fischer, Mehrad Moradshahi, Silei Xu, Jackie Yang, Richard Yang, and Monica S. Lam. Almond: The open, privacy-preserving virtual assistant. `https://almond.stanford.edu`, 2019.

[3] Giovanni Campagna, Rakesh Ramesh, Silei Xu, Michael Fischer, and Monica S. Lam. Almond: The architecture of an open, crowdsourced, privacy-preserving, programmable virtual assistant. In *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, pages 341–350, New York, New York, USA, 2017. ACM Press.

[4] Giovanni Campagna, Silei Xu, Mehrad Moradshahi, Richard Socher, and Monica S. Lam. Genie: a generator of natural language semantic parsers for virtual assistant commands. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*, pages 394–410. ACM, 2019.

[5] Giovanni Campagna, Silei Xu, Rakesh Ramesh, Michael Fischer, and Monica S. Lam. Controlling fine-grain sharing in natural language with a virtual assistant. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–28, sep 2018.

[6] Rich Caruana. Multitask learning. In Sebastian Thrun and Lorien Pratt, editors, *Learning to Learn*, pages 95–133. Springer US, 1998.

[7] Michael Fischer, Giovanni Campagna, Silei Xu, and Monica S. Lam. Brassau: Automatically generating graphical user interfaces for virtual assistants. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2018)*, 2018.

[8] Ranjay Krishna, Michael Bernstein, and Li Fei-Fei. Information maximizing visual question generation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[9] Douglas MacMillan. Amazon says it has over 10,000 employees working on Alexa, Echo. *The Wall Street Journal*, 2018.

[10] David Mazières. The stellar consensus protocol: A federated model for internet-level consensus. *Stellar Development Foundation*, 2017.

[11] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*, 2018.