Capstone Project Proposal

Tibor Zahorecz

Business Goals

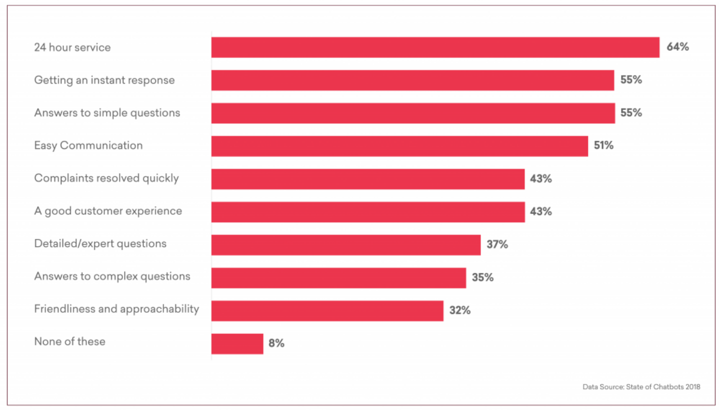
Project Overview and Goal What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?	**Industry problem definition**: A major problem for technology product vendors is after-sales (and pre-sales) support, that is, when a customer who purchases a complex technical product has a problem with the device or the software (it does not work as expected). In this case, the technical support division of the company becomes very important to avoid losing clients (called churn). However, building such a 24/7 service from such highly trained IT professionals is very expensive (or not possible due to missing IT experts on the field). **Goal definition**: we are helping product vendors to build responsive and economical after-sales support organization. **How does AI add value?** (Why use ML/AI in solving this task? To resolve the problem detailed above we can do: a. Add more Human Resources. Advantage: very adaptive, creative, human. Disadvantage: high cost, it can be scaled slowly b. FAQs, how-to videos on the company website. Advantage: Scalable, low update cost. Disadvantage: static, only certain types of people are inclined to read manuals, not responsive c., **Intelligent AI bot**. Advantage: scalable, adaptive, cheap Disadvantage: need very good input data, need monitoring in training and deployment period The solution is a mix of the above three items.

	How does AI add value? **Give "human type", adaptive support available when no Human Expert is available. Makes tech companies grow scalable.** **Be as specific as you can when describing how ML/AI can provide value** The AI bot should provide: 1. Level 1 Technical Support (this is a detailed semantic extract of the company FAQ table) 2. Point out the correct answer in the user manual of the product and extract to clients 3. Escalating the case of a serious hardware and / or a software tickets to human expert 4. Communication must be "human type" **Business benefits** 1. 24/7 availability for smaller (SME-sized) companies ==> reducing customer churn rate 2. Polite, professional service for Level 1 technical questions (these are often boring ones to professional IT supporters, so their style is characterized by tech arrogance) ==> increased customer experience **3.** It takes 6 months to train an AI bot on proper database and FAQ tables, but after then the workload can be scaled ==> cost optimization, scalable growth
Business Case Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.	ABC corp., with its Central European headquarter, manufactures high quality embedded products with AI components. The company's product line has high prestige in the market thanks to the image quality and software flexibility. But users need to have certain IT capabilities and have got some technical questions in the pre- and after sales process. **The company cannot attract enough IT staff** because there is an over-demand in the region for people with this education background. Another issue is IT people refuse to work in two or three shifts (24/7 market expectation). In the short term, the company seeks to remedy the situation by continuing to train its sales colleagues technically and involve them in the pre- and after sales processes. Which led to sales overload causing high level of fluctuation. Ownership expectations are high, so the company **management is willing to invest in internal IT** developments. In this case, **an automated, highly trained AI online technical support chatbot would have a huge business impact** solving sales and technical people overload.

Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	**Process of building chatbot service** • 1st step learning on the company Tech support conversations (Ticketing system) • 2nd step Annotation Team (company Tech support team & dedicated experts) train the bot, review the bot answers • 3rd step Automate the chatbot on the company website **What precise task will you use ML/AI to accomplish?** 1. Chatbot ensuring consistent customer support experience. Chatbot is available 24/7 to help customers with level 1 issues, FAQs, manuals. (MVP phase) 2. Integrate the chatbot with company CRM system to send customer queries to human Technical Support for further assistance (escalation). (Post-MVP phase) 3. Show customers the route for solving technical problems online. (MVP phase) 4. Chatbot is designed with troubleshooting workflows. Guide customers to find solutions and encourage independent problem solving. (MVP phase) 5. Ticket categorization. Chatbots simplify the process by intelligently categorizing tickets to the agent group based on the type of issue. (Post-MVP phase) 6. Knowledge Management & Know-how transfer is the key to efficient self-service. It helps largely in ticket deflection and is a great source of self-help. Chatbots are useful in suggesting the right articles and getting a real-time feedback. Chatbots auto-suggest relevant articles to the users. They identify knowledge gap based on the aggregated ticket data and identify missing knowledge articles. Based on the ticket resolution provided, new knowledge articles are created. (Post-MVP phase) **What business outcome or objective will you achieve?** • **24/7 Availability –**. This results in the overall improvement of resolution time and productivity. • **Faster ticket resolution** • **Proactive engagement –** Proactive support saves cost by eliminating delays and major service outage.

- **Cost optimization** – Chatbots deflect routine Level I tickets which ultimately reduces the overall after-sales support cost incurred.

- **Customer Experience**



Data source: State of Chatbots 2018

Success Metrics

Success Metrics

What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.

General guidelines:
Monitor the accuracy, the performance and fairness of your AI model and understand the reasoning behind the results. But focus on the outcome and not the output!
Outcome: revenue generation, improve customer experience, increase user satisfaction automate business operation and cut costs.
Output: accuracy, execution time, recall, precision

Business metrics (outcome)

Nr. of active users vs resolved level 1 tickets

User Interactions with the bot
The number of interactions that each user has with our bot

Average chat sessions

Goal Completion Rate (GCR)
Chatbot is built in such a way that it can answer general technical questions and increase customer experience (net promoter score).

Resolved level 1 tickets by chatbot

Customer satisfaction level with chatbot

Baseline: Current Ticketing system statistics, that is human experts' results: resolution time, average sessions vs tickets, customer satisfaction level.

ML metrics (output)

Confusion Matrix

True Positive (TP): • Reality: A wolf threatened. • Shepherd said: "Wolf." • Outcome: Shepherd is a hero.	**False Positive (FP):** • Reality: No wolf threatened. • Shepherd said: "Wolf." • Outcome: Villagers are angry at shepherd for waking them up.
False Negative (FN): • Reality: A wolf threatened. • Shepherd said: "No wolf." • Outcome: The wolf ate all the sheep.	**True Negative (TN):** • Reality: No wolf threatened. • Shepherd said: "No wolf." • Outcome: Everyone is fine.

source: Google developer

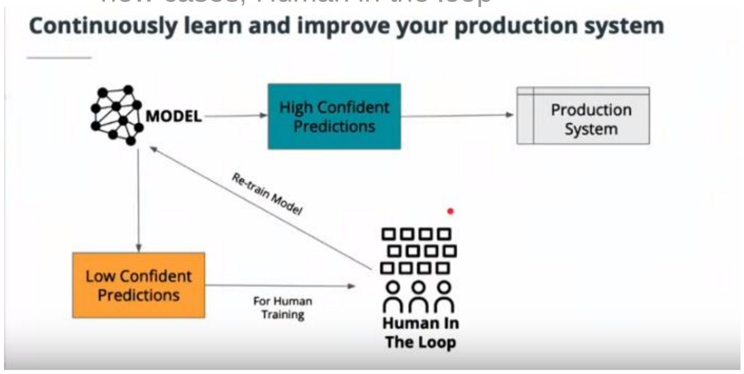
	Precision Precision identifies the frequency of correct answers, when the prediction is intent A. It can be thought of as the answer to the question “Out of all predictions of A, how many were correct?” **Recall** Recall identifies the frequency of detecting A, out of all examples pertaining to A in reality. In short, it answers the question “out of all the examples in A, how many were detected?” **F1-Score** F1-Score calculates the harmonic mean of precision and recall. It helps you answer the question “What is the global performance of prediction, with respect to class A?” **Accuracy** Accuracy is often the go-to metric to measure performance. It is the fraction of all predictions that are correct.

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	**Where will you source your data from?** From the SQL database of ABC corp. technical support ticketing system. ABC corp product's User manual, Installation Guide, FAQ lists, Programming Manuals (in English). Internal data which ABC Inc's property. **What is the cost to acquire these data?** As these are internal databases and structured Text format, there is no extra cost to acquire. Additional costs coming from data cleaning process. **Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome?** The SQL database contains the names of the notifiers (not the company name) and the e-mail addresses. These fields are not used in bot teaching. **Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?** The database and text files mentioned above are constantly evolving thanks to the continuous work of human technical experts and product managers. That is, the AI bot can learn and improve continuously.
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	**Bias types:** Model bias, Data bias, Annotation bias **What biases are built into the data?** The current English technical support database (SQL based) has been built over the past several years by about 10-12 human experts. However, the depth and semantic diversity is good enough to train the chatbot but include some bias: 1., English correctness: as it is written by non-native English speakers, it has a big difference in word and sentence usage 2., Error categorization (mislabeling): The categorization of the incoming tickets sometimes is not obvious and clear 3., Differences in skill levels, i.e. the standard deviation of response quality 4., Cultural and gender bias: Since the technical support team is a white male, a Central European race, there are biases towards other races and genders in the database

<table>
<tr>
<td></td>
<td>
How might the data be improved?
We need diversity where subjective opinions matter.
Solving bias has a process:

1st Awareness: defining the decisions you are asking your model to solve for, is the decision is simple? (black and what or…),
2nd Data: data source, cover enough edge cases? Think about end users and test with them, ➔ gather diverse ML group to ask diverse questions
3rd Iterate & Learn: end user feedback, learning new cases, Human in the loop



source: Udacity AI product mgmt nanodegree course

Identifying the most valuable training data units for human annotation like class importance, uncertainty, low confidence.
Annotation Team (company Tech support team and business unit mgmt) review the bot answers, perfecting its answers and then is tested with a small audience.
</td>
</tr>
<tr>
<td>Choice of Data Labels
What labels did you decide to add to your data? And why did you decide on these labels versus any other option?</td>
<td>
What labels did you decide to add to your data?
Hardware problem
Software problem
OS problem
AI engine problem
Documentation
Dead on Arrival
Why did you decide on these labels versus any other option?
The clusters of the current database error categorization make these labels the minimum necessary.
</td>
</tr>
</table>

Model

Model Building

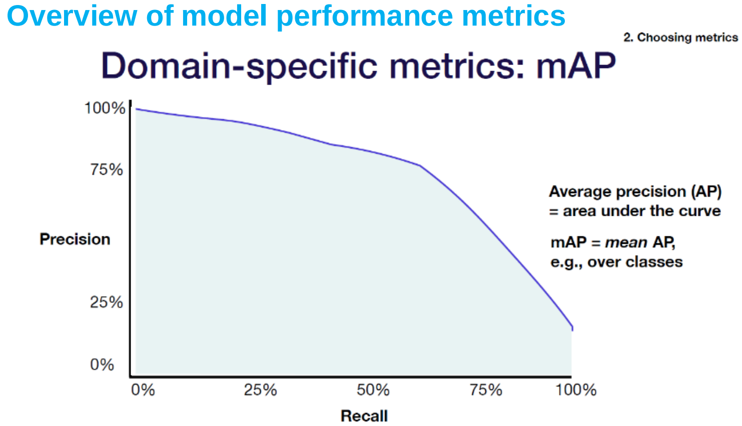
How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?

This step is closely related to MVP and rollout sections. Since successful internal enterprise development requires the approval of internal stakeholders (CEO, CMO, CTO etc) and keeping costs to a minimum level, so model development needs to be broken down into two major steps.
1., Demonstrate and prove to internal stakeholders that a high quality technical virtual agent can be achieved and create value to the organization. This should be done at minimal cost on an existing platform provider under an MVP development: i.e. IBM, Google, Microsoft (see MVP section for specific example) cloud platforms.
2., If the internal stakeholders engage in the project then a technology feasibility study should be done: cost, teams, internal or external, technology etc. to decide inhouse development or using platforms with API.

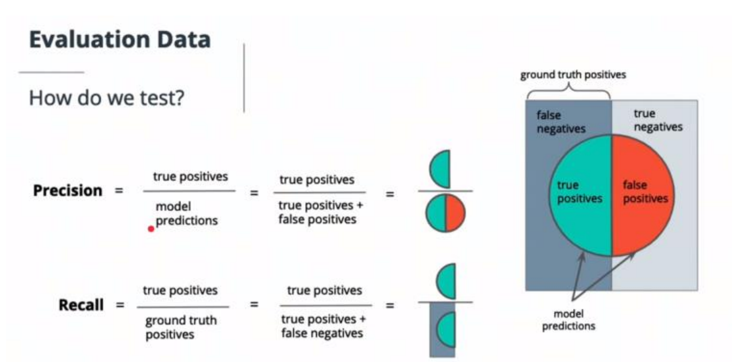
Evaluating Results

Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?

Overview of model performance metrics



source: https://fullstackdeeplearning.com/march2019



source: Udacity AI product mgmt nanodegree course material

Which model performance metrics are appropriate to measure the success of your model?

I use the following machine learning metrics: Precision, Recall, F1

What level of performance is required?

Defining a good performance metric assumes a well-chosen baseline. This requires industry benchmark and research papers. From my research, I have taken the following document and metrics as a baseline.

Evaluating Natural Language Understanding Services for Conversational Question Answering Systems: https://www.aclweb.org/anthology/W17-5522.pdf

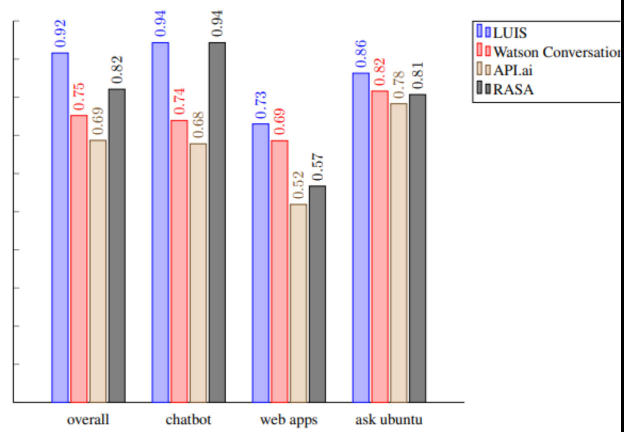
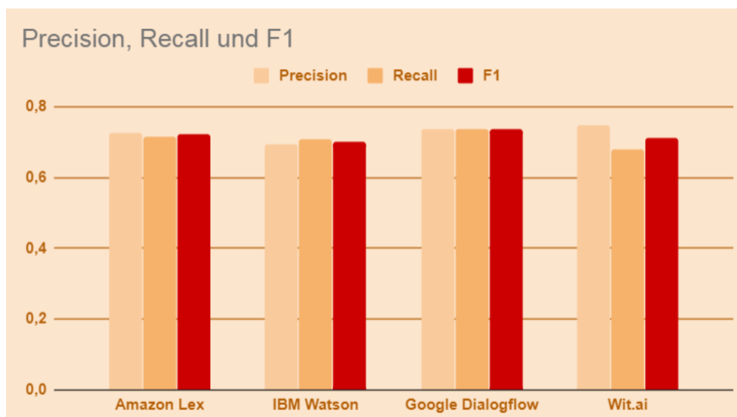


Figure 3: F-scores for the different NLU services, grouped by corpus



Another good source of baseline: https://medium.com/analytics-vidhya/quality-metrics-for-nlu-chatbot-training-data-part-1-confusion-matrix-91ac71b90270

	** background information for model performance metrics **What are precision and recall?** Precision and recall help us understand how well our model is capturing information, and how much it's leaving out. Precision tells us, from all the test examples that were assigned a label, how many actually were supposed to be categorized with that label. Recall tells us, from all the test examples that should have had the label assigned, how many were actually assigned the label. A high precision model produces fewer false positives. A high recall model produces fewer false negatives. **F1 Score** combines precision and recall into a single number, which makes comparing two models easier. $F1 = \frac{2 * Precision * Recall}{(Precision + Recall)}$

Minimum Viable Product (MVP)

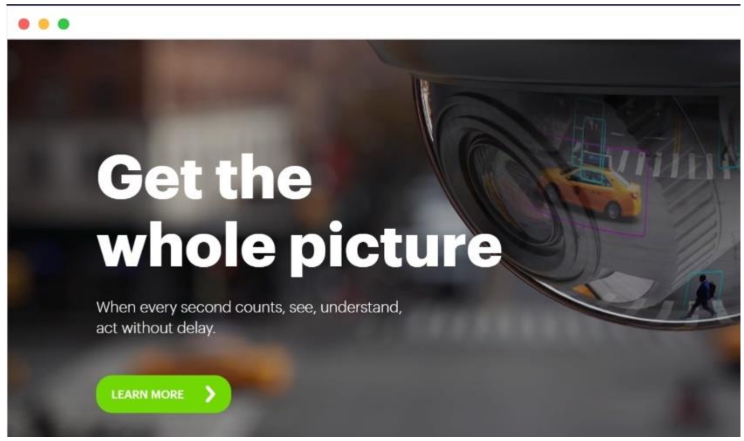
<table>
<tr>
<td>

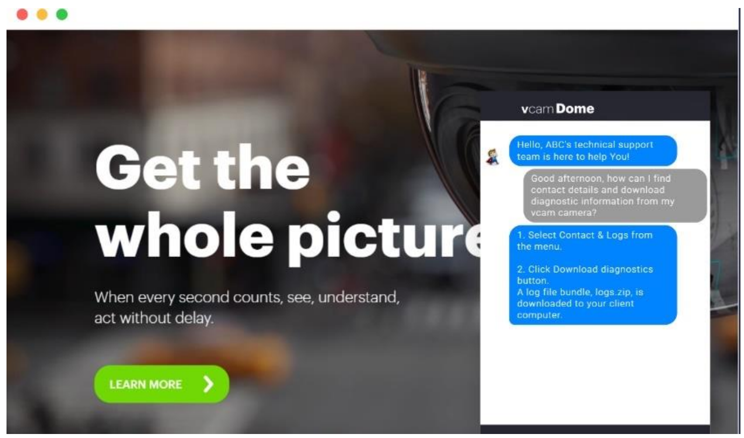
Design

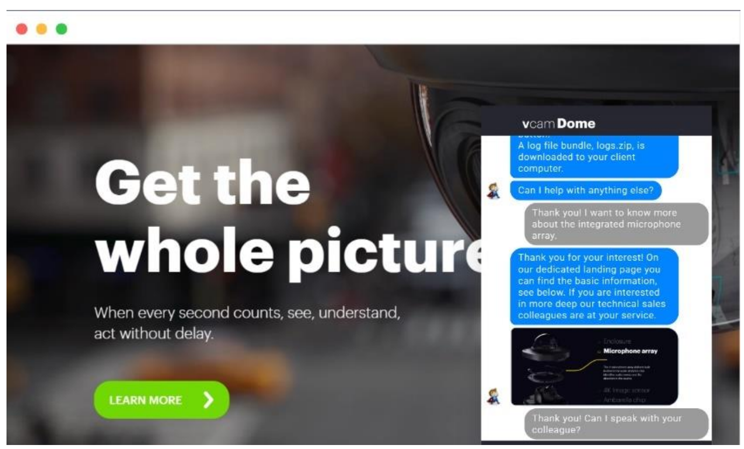
What does your minimum viable product look like? Include sketches of your product.

</td>
<td>

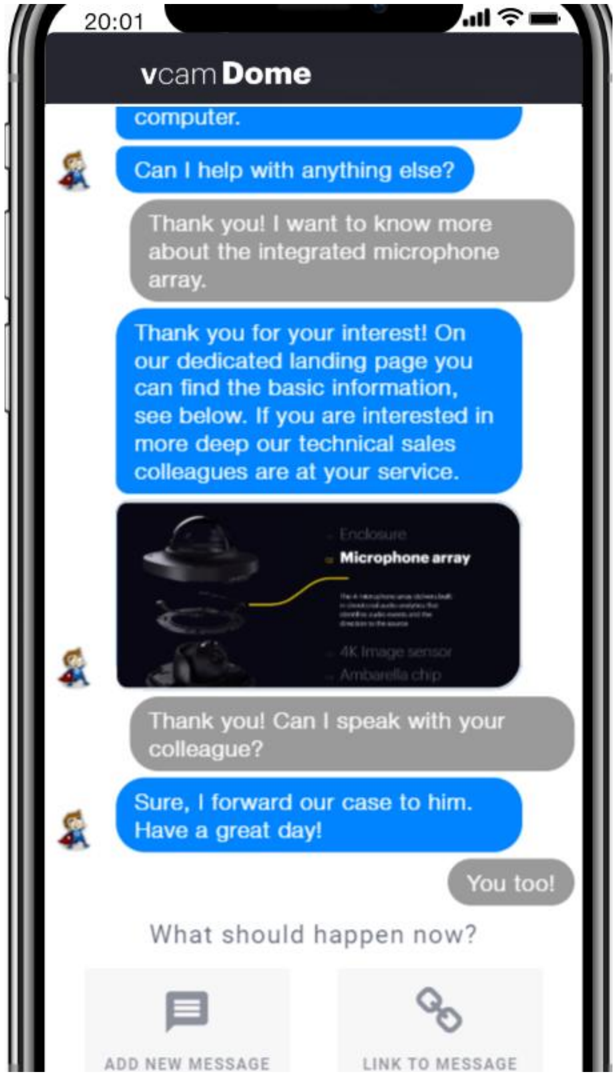
MVP sketches was made upon: https://botsociety.io/





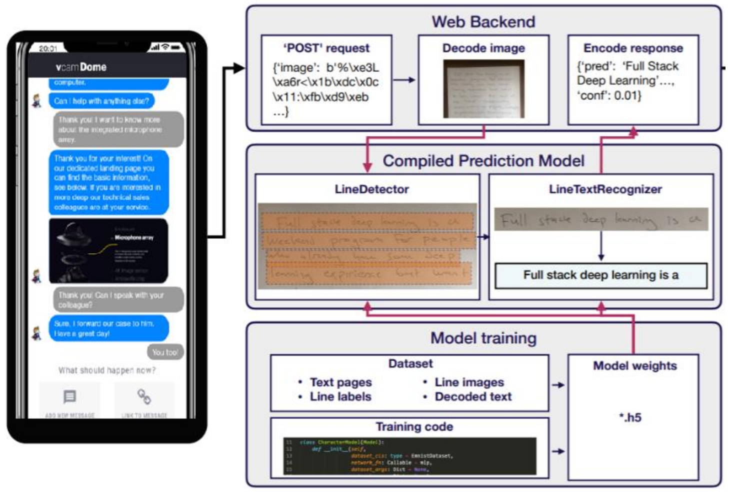


</td>
</tr>
</table>



You can check and edit design here:
https://app.botsociety.io/s/5e2355f09a190468ac02ef94?p=9b73f2505b11621389430b951363453aa2a86340&desktop=true
(attached the sketches in the zip file)

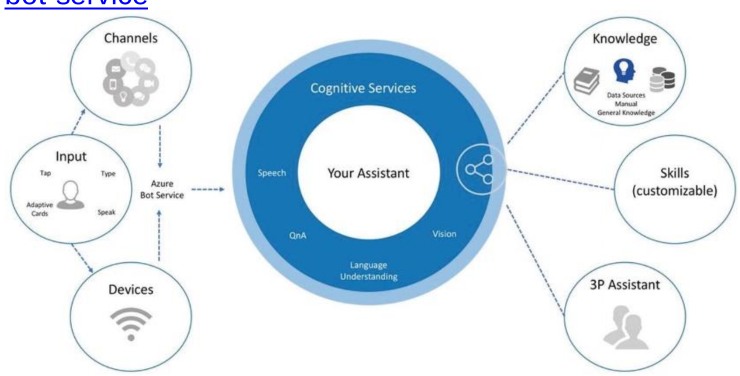
Architecture
I follow Fullstackdeeplearning bootcamp demo application development
https://fullstackdeeplearning.com/march2019

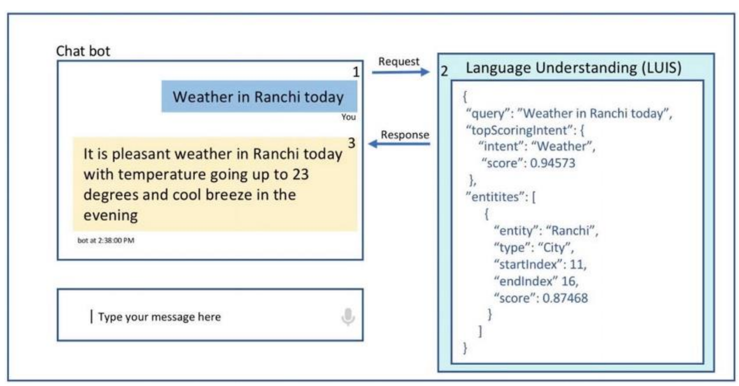


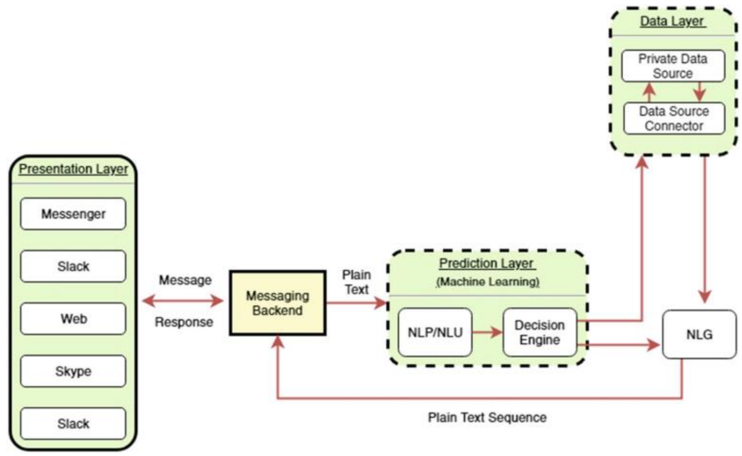
Technical MVP

The technical MVP demonstrates the operational logic and feasibility of the chatbot service to the project's main stakeholders.

a., Microsoft: How to Build A Chatbot FAQ With The Azure Bot Service (https://www.qnamaker.ai/) https://docs.microsoft.com/hu-hu/archive/blogs/mvpawardprogram/chatbot-faq-azure-bot-service



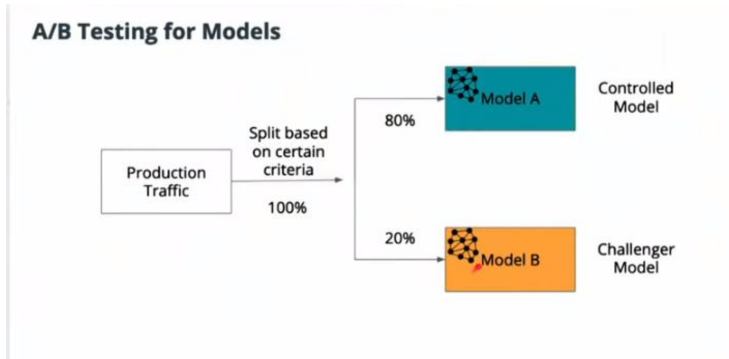


	b., IBM Platform  https://developer.ibm.com/patterns/compose-bots-using-an-agent-bot/ https://github.com/IBM/support-ticket-classification
Use Cases What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?	The chatbot's user base will be expanded after the MVP phase, as I will explain in the post-MVP phase. **User Profiles and Use Cases in the MVP Stage**: A., In-house Technical Support colleagues alias Human Expert, Product Managers (supervisor), Sales Managers (supervisor). Use Case: supervising, teaching the bot service. B., The MVP chatbot service will be under the company website/portal. Only one product feature will be available at the beginning of the MVP phase that is level 1 technical FAQ. **Use Case**: Level 1 technical Q&A for hardware, software and AI engine issues. Level 2 issues (like software code explanation) will be escalated to the Technical Support Team (Human Experts). **User Personas**: Users of the products with minimal IT technical knowledge, not reading user manuals and written FAQs on the portal as not problem solvers types. Example: Mike is working as electrical technician at Woldborg ltd integrator company. The company itself doing bank IT system integration. Mike is working on the field with security cameras. He has no software or IT background. He's very frustrated when camera integration doesn't go

	through standard interfaces because he doesn't understand new software technologies, but he doesn't want his employer to find out. The camera manufacturer's user guide or software guide doesn't help him.
Roll-out How will this be adopted? What does the go-to-market plan look like?	The Technical Support chatbot is placed on the product support page of ABC corp. Portal. Go-to-market plan: After internal iterations, when the stakeholder team is convinced of the technical readiness of the chatbot, it goes to the company website. In the first phase (MVP phase), the client has the option of manually opening support ticket after the bot responses and give feedbacks its performance.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	**Product will be improved in the long-term** • Using product feedback/new information from users and applying the feedback to new product versions being released • Continuous repeat of prototyping and iterating process quickly and faster with new user information • Refreshing and updating the model with new data to prevent model staleness where predictive power of the ML model decreases over time **Real-world data is different from the training data** • Training data is used when a model is under learning phase • Real-world data obtained from various real-world sources that are associated with outcomes of chatbot performance **Product can learn from new data** • Setting up the model in a dynamic way that it is continuously trained from new input • Updating the data to include more relevant examples for model to gain expertise • Getting rid of incomplete and inaccurate data • Ensure that training and test data are balanced **Future Product plan (post-MVP)** We extend our chatbot service to learn on the product manuals and programming guides. We will create an internal service based upon the company MS Dynamics environment. Target is to build a Virtual Assistant for our internal training programs to reinforce our new sales and representatives onboarding program. At a later stage, the product will evolve into an extended pre-sales support to assist with sales negotiations. ** Dynamics 365 Virtual Agent example https://poszytek.eu/en/microsoft-en/dynamics-365-virtual-agent-2/ **A/B testing** helps to make more data-driven decisions Deploying two different chatbot models. Splitting acquired data into two models in order to track performance metrics associated to each model; after which a decision can be made on the winning model and/or if the lowest performing model can be replaced

	with the better performing winning model. Designing A/B test for Models • performance metric decision • choosing a minimum effect size, we want to detect • determining the sample size • running the test until sample size is reached  source: Udacity AI product manager nanodegree program **What else we should consider?** • Cost benefit analysis a., is x% accuracy gain beneficial for our business? b., what if this slightly better model requires a much larger investment? • Run the test long enough to capture any seasonality effects • Control the experiment to avoid “novelty effect”
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	**Monitoring and mitigating bias** should be an ongoing task. Models are going to be as good as the data we provide. If we add bias (wrong labels, unbalanced labels, etc) we will have an imprecise or useless model. Practically, we are always going to have some kind of “bias”; what we want is lower it until an acceptable point. Data: the more data, the better. Particularly, for cases hard-enough to label. In our case we need data from out gender (women) and culture (China, Spore, Niger). Iteration and improve. We add these new data to our learning model.