

# Profile of Aiaioo Labs

**Cohan Sujay Carlos**

Aiaioo Labs

Bangalore, India

<http://www.aiaioo.com>

[cohan@aiaioo.com](mailto:cohan@aiaioo.com)

## Abstract

The constraints (financial and operational) under which a research SME operates make its goals and strategies very different from that of a product firm's research team. In this article, we describe Aiaioo Labs, a research SME in Bangalore. We take a look at what Aiaioo Labs does, and the reason it is valuable in the backdrop of the Indian research environment.

## 1 Introduction

Aiaioo Labs was founded to provide research services in applied AI to firms in and around Bangalore. The goals of the lab were as follows:

- Decrease research costs to local industry.
- Decrease technology risks to clients.
- Increase the availability of special skills.

We help firms decrease the cost of doing research when they partner with us by supporting them with tools, components, data, architectural assistance, training and research assistance.

We help decrease technology risks to client firms by maintaining a set of ongoing research projects that push the technology in those verticals.

We help encourage the development of research skills by providing students interested in research with a place where they can get paid to do fun things and also meet and work with like-minded people.

## 2 Strategy in India

A research SME's strategy is determined in large part by local economic conditions. The market opportunities available to local entrepreneurial ventures and the state of the ecosystem that supports

high technology product development determine to a large extent the strategies of a research SME.

The main components of an ecosystem that supports high technology product development are:

- Risk Takers
- Institutions of Higher Education
- Large Firms
- Government Support

In India, the last three ingredients are not readily available to the entrepreneurial ecosystem for the following reasons:

- Indian universities produce very few science PhDs compared to other developing and developed countries.
- Most of the work coming to India is in the services industry.
- Research grants are not available to private organizations.

So our team works not just on developing algorithms (research), but also on applications of those algorithms (productization), and on offerings that solve business problems (solutions).

The rest of this document discusses our research projects, our patents and our publications.

### 2.1 Compiler Front-End

Our first publication as a lab in 2011 concerned a system for Natural Language Programming that allows users to program computers by issuing commands in a non-formal language. The publications on the topic and a demo can be viewed at <http://www.aiaioo.com/publications>.

## 2.2 Mass Communication Tool

A Mass Communication Tool (MCT) is one that can support a live conversation between a few people and a large number of people. Such a conversation would require a mass communication system that uses automation in the form of natural language processing and machine learning tools to enable a few people of note (for example, the CEO of a company) to interact with a large number of people (for example, all the employees of the company) in real time (for example, during a virtual all-hands meeting), get a synopsis of a large number of concurrent discussions in real time, and participate in a significant fraction of the discussions as they are taking place.

The system would consist of:

- an aggregator of messages (built from natural language processing components) that groups together messages and discussions with identical semantic content;
- a hierarchical clustering system (built from natural language processing components) that assigns aggregated messages their place in a hierarchy by specificity with more general messages closer to the root of the hierarchy and more specific messages closer to the leaves of the hierarchy;
- a summarization system (built from natural language processing components) that creates a summary of the aggregate of all messages in a sub-tree; and
- a reply routing system (built from natural language processing components) that routes replies from a few people of note to their appropriate places in discussion threads.

The operation of the system would not always have to be dependent on AI tools. It would be possible to do some of the work through crowd-sourcing, for example:

- the aggregation of messages can be assisted by inputs obtained from one or more users;
- the hierarchical clustering can be assisted by inputs obtained from one or more users;
- the routing of replies can be assisted by inputs obtained from one or more users.

MCTs are bound to have a huge impact on our experience of representative government. A typical use case would involve a public figure, (say President Obama), sounding out the electorate before introducing legislation on say healthcare reform.

By first discussing the competing proposals with large numbers of people, it might be possible for the initiator of the discussion to get a sense of what might or might not work and what the response to the legislation was likely to be.

Apart from decreasing corruption by giving honest politicians a way to showcase their abilities, the tool has the potential to sustain a process of continuous involvement of people in government. The role of the electorate is now no longer limited to its participation in elections once every few years.

It seems very possible that identifying corruption is not necessary for reducing corruption. What seems more likely to work in the long term is not policing mechanisms for identifying corrupt individuals, but the creation of conditions within the system that allow honest individuals to eventually become more competitive than dishonest individuals owing to improved transparency (possibly through dialog with the electorate).

In the seminal paper, ‘The Market for “Lemons”: Quality Uncertainty and the Market Mechanism’, George A. Akerlof explained that quality uncertainty led to bad cars driving out good cars, in the same way that bad coins drive out good coins from circulation.

A similar effect might exist in politics since corrupt candidates would have more resources at their disposal and thus be able to out-spend more honest candidates for public office. The honest candidates would therefore be at a considerable disadvantage unless there existed a mechanism that helped them prove their quality.

In the used car market, something that could act as an indication of quality is an inspection report from a trusted mechanic. Similar mechanisms for signalling quality are what would be needed to improve the chances of honest people getting elected and entering influential positions in politics.

MCTs might serve the purpose and help act as quality beacons for honest politicians.

### 3 Research in Progress

Tools for Business Intelligence offer users a fairly standard set of analytics capabilities that include, among others, classification, clustering, association rule mining, OLAP cubes and query languages for the same. All of these capabilities are focused on structured text that usually originates in what is known as an OLTP system. Thus there is a need for a strategy for integration of unstructured data in the BI pipeline. This is particularly relevant in the light of a Gartner report that claims that unstructured information comprises no less than 80% of the useful information available to a company.

Our immediate research work focuses on a method of text analytics that allows commercially valuable forms of textual utterances to be converted into structured tables that can be stored in an RDBMS (also known as the OLTP system). This structured information can be analysed using any of the available BI analysis tools - using OLAP cubes for instance. The conversion of the unstructured information into structured form also makes it possible to use the standard reporting and visualization tools of BI products to display results of analyses to users (in the form of standard graphs and charts).

The key to this has been the development of automated methods for the extraction of intention from utterances. Intentional utterances are an old concept that date back to the 1960s (they were studied by Searle and Winograd as part of the linguistic theory of Speech Acts). However, there has been very little subsequent work aimed at automatically identifying Speech Acts in text.

We are in the process of curating a corpus for studying intention and there is a whitepaper available on our website.

Below are some of our other projects at different stages of completion.

#### 3.1 Graph Analysis

We conducted a study of possible methods of analysing communication networks to identify participants in terrorist attacks. We found out that for multiple simultaneous attacks on a city the size of Mumbai, it would be possible, if the attackers had coordinated their actions using the cell phone network, to identify the participants with very high accuracy, if three or more locations were attacked by the same group of attackers.

	Resolve	Ignore
Social Channel	4, 3	-2, -5
CRM Channel	4, -2	-2, 0

Table 1: Sample Game Theoretic Model

#### 3.2 Game Theory

Some of our recent work focuses on the use of game theory to model the delivery of services. We are attempting to model a firm's incentive to respond to complaints online and customers' incentives to use online channels to complain.

The opposing motivations seem possible to model as a static game with complete information where there are two players, the aggrieved customer, and the firm, as shown in Table 1.

Our investigations so far suggest that the use of social media for customer service might benefit from an understanding of the game theoretic motivations underlying customers' and firms' choices.

### 4 Patents

We have taken out patents on the following:

- Temperature-controlled clothing that can be produced at low cost.
- Compiler Front-End using Natural Language Processing Tools.
- A mass communication tool using Natural Language Processing.

#### Acknowledgements

This document was prepared in the month of February, in the year 2012 with help from our researchers ... so many thanks to:

- Sumukh - for innumerable edits and tips.
- Madhulika - for being part of the team.