**Investigating the team planning strategies of the human-AI teams in a bomb disposal game**

Lixiao Huang, Prithvi Prathapan, Sai Sushank Manthri Pramod

**IFT 536**

**Spring 2024**

**Table of Contents**

[**Abstract (Lixiao) 3**](#_heading=h.tyjcwt)

[**Introduction (Sushank, Lixiao, Pri) 3**](#_heading=h.4i7ojhp)

[**Team Planning** 3](#_heading=h.2xcytpi)

[**Artificial Social Intelligence for Successful Teams (ASIST)** 4](#_heading=h.1ci93xb)

[**NLP Principles and Theories** 4](#_heading=h.3whwml4)

[**Research Questions** 5](#_heading=h.2bn6wsx)

[**Method 5**](#_heading=h.qsh70q)

[**Experimental Design (Lixiao)** 5](#_heading=h.3as4poj)

[**Participants (Lixiao)** 5](#_heading=h.1pxezwc)

[**Dataset (Lixiao)** 6](#_heading=h.49x2ik5)

[**Data Analysis** 6](#_heading=h.2p2csry)

[***1. Manual coding approach:*** *6*](#_heading=h.147n2zr)

[***2. The AntConc approach*** *8*](#_heading=h.3o7alnk)

[***3. The machine learning approach*** *9*](#_heading=h.23ckvvd)

[**Results 9**](#_heading=h.ihv636)

[**Manual Coding Results:** 9](#_heading=h.32hioqz)

[**AntConc Results:** 9](#_heading=h.1hmsyys)

[***Top 20 words by condition*** *9*](#_heading=h.65nn3goahros)

[***Wordcloud results:*** *11*](#_heading=h.41mghml)

[**Machine Learning Approach Results** 12](#_heading=h.2grqrue)

[**Discussion 13**](#_heading=h.vx1227)

[**Data Issues** 13](#_heading=h.3fwokq0)

[**Insights** 13](#_heading=h.1v1yuxt)

[**Limitations** 14](#_heading=h.4f1mdlm)

[**Future Directions** 14](#_heading=h.2u6wntf)

[**Appendices 14**](#_heading=h.19c6y18)

[**Appendix A: Stop List** 14](#_heading=h.3tbugp1)

[**Related References 2**](#_heading=h.28h4qwu)

**Investigating the team planning strategies of the human-AI teams in a bomb disposal game**

Lixiao Huang, Prithvi Prathapan, Sai Sushank Manthri Pramod

**Abstract**

Team planning is essential for mission success in a human-AI teaming task environment. This project investigated the human team planning quality through their text chat communication and its correlation with team outcomes. We used three approaches to analyze the data: manual code, AntConc, and machine learning. The manual coding approach gave us the best coverage of the content themes but was inefficient in coding large datasets. AntConc is helpful in showing word frequency in the whole dataset. The machine learning approach to classification is also fast but has low reliability. In the future, we plan to use the existing results to train a better model and calculate the correlations between the found identified patterns and the team scores.

**Introduction**

**Team Planning**

Team planning is essential for mission success because it impacts the team strategies and execution, especially for complex task scenarios. Team planning is a vital process within teams that involves laying out a course of action to achieve predefined objectives. It's a proactive approach that enables teams to coordinate activities effectively, particularly in complex decision-making tasks requiring analytical skills. Team planning applies to healthcare, battlefield, and other team tasks (Is there a reference for this definition?).

Evaluating the quality of team planning is the key to achieving high-quality team planning. Measuring team planning typically involves assessing how teams engage in systematic and strategic planning processes (Oldeweme et al., 2021). This can include evaluating the clarity and specificity of goals set by the team, the comprehensiveness of the plans developed, the allocation of tasks and responsibilities among team members, and the consideration of potential obstacles and contingencies (Oldeweme et al., 2021).

Major challenges in team planning involve effective communication, resource allocation, task allocation, and strategies for in-game coordination. Best practices in team planning emphasize the importance of collaboration, communication, and flexibility. Teams benefit from involving all members in the planning process to ensure diverse perspectives are considered and consensus is reached. Clear and specific goals should be established, with tasks and responsibilities clearly defined to avoid confusion and duplication of effort. Regular review and adaptation of plans based on changing circumstances or new information are essential to ensure the team remains agile and responsive to challenges.

**Artificial Social Intelligence for Successful Teams (ASIST)**

The goal of the ASIST project is to develop an AI agent capable of inferring human team members’ states, predicting their next actions, and intervening to improve the team’s effectiveness during planning and action. For the AI agent to give effective interventions, the team’s planning quality must first be evaluated. Team planning is mainly reflected through their text chat messages at the shop.

**NLP Principles and Theories**

Natural Language Processing (NLP) is a field of study at the intersection of artificial intelligence, linguistics, and computer science. It aims to enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful. Here are some fundamental principles and theories of NLP:

1. Linguistic Theory: Linguistic theories provide the foundation for understanding the structure and meaning of human language. They include syntax (the structure of sentences), semantics (the meaning of words and sentences), and pragmatics (the context in which language is used).
2. Tokenization: Tokenization is the process of breaking text into smaller units, such as words or sentences. It's a crucial step in NLP tasks as it forms the basis for further analysis.
3. Part-of-Speech (POS) Tagging: POS tagging involves labeling each word in a sentence with its corresponding part of speech (e.g., noun, verb, adjective). This information is essential for understanding the grammatical structure of sentences.
4. Named Entity Recognition (NER): NER is the task of identifying and classifying named entities (e.g., persons, organizations, locations) within text. It's vital for information extraction tasks.
5. Syntax and Parsing: Syntax refers to the grammatical structure of sentences, and parsing involves analyzing the syntactic structure to understand relationships between words and phrases. Dependency parsing and constituency parsing are common techniques used in NLP for this purpose.
6. Semantic Analysis: Semantic analysis focuses on understanding the meaning of text beyond its grammatical structure. Techniques such as word embeddings, semantic role labeling, and sentiment analysis fall under this category.
7. Machine Learning and Deep Learning: Machine learning and deep learning techniques play a significant role in NLP, enabling computers to learn patterns and relationships in language data. Models like recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based architectures (e.g., BERT, GPT) have shown remarkable performance in various NLP tasks.
8. Language Models: Language models are statistical models that learn the probability of sequences of words. They are used for tasks such as language generation, machine translation, and text summarization.
9. Discourse Analysis: Discourse analysis focuses on understanding how language is used in context, including the structure of conversations, coherence, and cohesion.
10. Pragmatics: Pragmatics deals with the study of language in use and the context in which communication takes place. It considers factors such as speaker intentions, implicature, and conversational implicature.

These principles and theories form the basis for a wide range of NLP applications, including chatbots, virtual assistants, machine translation, sentiment analysis, and information retrieval. As the field advances, new theories and techniques emerge to address the complexities of human language understanding and generation.

**Research Questions**

This study aims to address the following research questions: (a) **Do people have different planning qualities in different AI advisor conditions? (b)** What themes do people talk about in their planning? (c) How frequently do people talk about these themes? (d) How do people’s talk patterns change over time? (e) How do people’s team planning relate to their team outcomes?

However, this final project for the class will focus on addressing questions (a) and (b) only.

**Method**

**Experimental Design**

This is a within-and-between-group mixed design of three advisor conditions (1 no advisor, two ASI advisors) applied to n trials of equivalent difficulty in a randomly selected order (see Table 1).

**Table 1.** Experimental Design

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Equivalent Mission Difficulty (within-team factor) | | |
| Mission 1 (trial 1) | Mission 2 (trial 2) | Mission n (trial n) |
| Advisor conditions (between-team factor) | No advisor (control group baseline) #1 | A total of 399 trials | | |
| CMURI ASI #3 | A total of 379 trials | | |
| DOLL ASI #5 | A total of 385 trials | | |

**Participants**

Participants were recruited from Arizona State University and social platforms (e.g., Reddit, Discord, etc.). Participants must be 18 years old, physically located in the US (international participants will require additional authorities to approve IRB), have reliable internet (We may not be able to verify it, but will try to log latency), have experience playing Minecraft with a standalone mouse and keyboard, speak English, have normal color vision. Minors will be excluded, as will prisoners. The recruitment will include participants who have participated in other 3-person Minecraft studies in our lab. Participants are allowed to come back and repeat participation. They can sign up with someone they know but must answer a question about whether they know each other outside the game.

**Dataset**

The anatomized data will come from the following dataset:

Lixiao Huang; Adam Fouse; Nancy Cooke; Edward Weiss, 2024, "Artificial Social Intelligence for Successful Teams (ASIST) Study 4 Dragon Testbed Dataset", https://doi.org/10.48349/ASU/ZO6XVR, ASU Library Research Data Repository, V2, UNF:6:jkhVIRagnIe25M/7ClVYUg== [fileUNF]

**Data Analysis**

In this project, we tried out three approaches: (a) Manual coding, (b) Using AntConc, (c) Using machine learning.

**1. Manual coding approach:**

We will start with manual coding by creating dimensions and categories based on the data we get. We will also define each dimension and provide raw data examples for them. As we code the data, if we encounter new items that cannot be covered by the code dimensions, we will refine the definitions by combining or separating the dimensions as needed.

1. Clean the data: We have 1160 data trials, totaling xx participants.
2. Manual coding to identify the themes

Three teams of each group = 9 teams

1. Develop the codebook to see what categories are out there. Table 2 illustrates the codebook for team planning.

Table 2 Planning Codebook

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Dimension | Explanation | Examples |
|  | priority | Coding as statements on strategy | - “important”  - “i need to do the bomb first” |
|  | request | Coding as statements requesting information or help, but not a yes and no type of confirmation | - "need 44”  - “need 41”  - “lwemme do it” |
|  | confirmation | Coding as giving a confirmative answer | “time works” |
|  | seek confirmation | Coding as simple binary questions | “time as in 7 min or ?”  “also continue team for 3 games ?”  “I thought you got that bravo?  yeah?”  “am i right” |
|  | command | Coding as commands and instructs for others to do certain things | “sure”  “ok tyy”  “works for me as long as we're fastt”  “yea”  “alright lets do this”  “cool”  “noice” |
|  | courtesy | Coding as greetings and polite expressions | “hi” |
|  | navigation | Coding as content that is about navigation strategies | “i take city?” |
|  | report | Coding as reports about individual and environmental status and situations | “5 minute marker”  “chain bomb”  “i'm stuck in A4”  “messed one up, rng on other  got 48”  “im done with 5 minute bombs”  “7 minyute mark”  “chained bomb” |
|  | goal setting | Coding as suggestions for goals | “need 36 done”  “i'm gonna get 48 after this one”  “I'll go to 44 and that other bobm”  “I'll grab that if can after 44” |
|  | inquiry | Coding as statements asking any questions to seek information; | “whats the p;an” |
|  | express emotions | Coding as expressions of emotions | “;-;”  “damn 1 blue up” |
|  | coordination | Coding as statements of coordinating actions | “ripill get you ih”  “ill get 41 no worries”  “Let me know when you got 35 so”  “I can compass”  “properly”  “okay, shop when 35 is done delta and I'll compass then”  “dont buy it” |
|  | seeking attention | Coding as signals for the listener to hold on momentarily before proceeding with their query | "wait" |
|  | correcting | Coding as correcting self or others’ statements | “how\*  you\*  and by delta I mrant alpha lol” |
|  | explanation | Coding as explaining a situation | “umm i clicked before i blew up” |
|  | seek clarification | Coding as statements that other members have mentioned earlier and then ask again | “37?” |
|  | other | Coding as phrases that are beyond comprehension | “il” |

1. Calculate the percentage of themes by trials; tally the frequency and calculate the total number of each, as well as by Agent conditions. - This step is not executed.
2. Compare groups to see the percentage patterns of themes. - This step is not executed.

**2. The AntConc approach**

AntConc is a free tool for corpus linguistics research (Anthony, 2023).

**Data Cleaning and Analysis**

Step 1: I started by cleaning the information's contents and getting rid of any extraneous or distracting knowledge, such HTML elements, grammar, and particular characters. After that, I used a stoplist to exclude terms that are often used but have no bearing on the analysis at hand.

Step 2: Apply stoplist. We found a stop list on the internet and saved it as a .txt file (see Appendix A). Then, go to the Global setting, under word list, click add a file, and upload the stoplist. Then, choose “hide the words.” Go back to the main menu and run the word tool again; the stopwords disappeared from the word list.

Step 3: Based on the word list without the stopwords, we generated a wordcloud (see results section).

Step 4: We generated the word list for all three conditions, took the top 20 words, and analyzed its pattern. We also tested out the KWIC, cluster, and n-gram tools. The frequency of the clusters and n-grams were not very high.

**3. The machine learning approach**

**Data Cleaning and Analysis**

Step 1: From the mixed list, select the first 1000 messages to use as training data for ChatGPT, and it resulted in 18 themes (collaboration - coordination - accidental - motivation - organizing - strategizing - encouragement - scheduling - clarification - disorganized - emergency - progress - warning - disengaged - efficiency - earthquakes - distracted - disruptive). Then use the 18 themes to predict the rest of the 11,000 messages.

In the machine learning approach undertaken, a mixed list of messages was subjected to a rigorous data analysis procedure. Initially, the first 1000 messages from this list were meticulously selected to serve as training data for ChatGPT. Through this process, 18 distinct themes emerged from the training dataset, encapsulating the underlying patterns and topics within the messages. These themes were then leveraged to predict and categorize the subsequent 11,000 messages. The data analysis involved comprehensive steps, including data cleaning to ensure consistency and accuracy in the training process. By filtering and preparing the dataset, the model was primed to effectively learn and extrapolate insights, ultimately contributing to the successful categorization and understanding of the larger message corpus.

**Results**

We present the results of the three approaches we have used: manual coding, AntConc, and machine learning.

**Manual Coding Results:**

The main results from the manual coding approach are the coding themes, including priority, request, confirmation, seek confirmation, command, courtesy, navigation

Report, goal setting, inquiry, express emotions, coordination, seeking attention

correcting, explanation, seeking clarification, and other.

**AntConc Results:**

We tried all 9 tools in AntConc, and selected the results that made the most sense to us.

**Top 20 words by condition**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| None | | CMU | | DOLL | |
| Type | Freq | Type | Freq | Type | Freq |
| can | 133 | ok | 122 | ok | 275 |
| forest | 110 | can | 118 | ready | 178 |
| ok | 110 | desert | 81 | bomb | 160 |
| one | 95 | bomb | 66 | go | 129 |
| desert | 91 | forest | 59 | get | 128 |
| bomb | 88 | im | 57 | one | 123 |
| s | 88 | s | 57 | forest | 122 |
| im | 83 | get | 56 | s | 120 |
| solo | 71 | ill | 56 | shop | 107 |
| get | 69 | just | 55 | desert | 106 |
| go | 65 | done | 54 | need | 104 |
| town | 64 | one | 53 | im | 93 |
| ill | 62 | town | 52 | got | 88 |
| need | 62 | go | 50 | bombs | 83 |
| got | 58 | got | 48 | buy | 82 |
| just | 57 | yeah | 42 | alpha | 76 |
| done | 56 | min | 41 | oh | 76 |
| yeah | 51 | solo | 40 | red | 73 |
| hi | 46 | time | 34 | green | 70 |
| ll | 43 | guys | 33 | ll | 70 |

In the top 20 words by advisor condition, None has five verbs: can, get, go, need, and got. CMU has four verbs: can, get, go, and got. DOLL has five verbs: go, get, need, got, and buy.

Surprisingly, DOLL condition did not have the word “can” on the list.

Get and go are the most common words in all three conditions’s top 3 verbs.

It also shows many teams did not talk much and the data quality is not great.

**Wordcloud results:**

The wordclouds below show that players talk about navigation a lot, such as desert, forest, town, and shop.

Wordcloud for the no advisor condition:



Wordcloud for the CMU advisor condition:



Wordcloud for the DOLL advisor condition:



**Machine Learning Approach Results**

The machine learning approach applied to the dataset yielded intriguing results, with an accuracy rate of 13% determined through rigorous evaluation. The accuracy was calculated by comparing the model's predictions against known outcomes within the dataset.

Despite the modest accuracy rate, this analysis provides valuable insights into the complexities of the dataset and the challenges inherent in predicting the themes of messages. The predictive model, trained on a subset of 1000 messages representing 18 distinct themes, demonstrated some capability in categorizing a larger set of 11,000 messages. Further refinement of the model and exploration of additional features or techniques may be warranted to enhance predictive performance and extract deeper insights from the dataset. Hence, building a predictive model was not useful with the low accuracy.

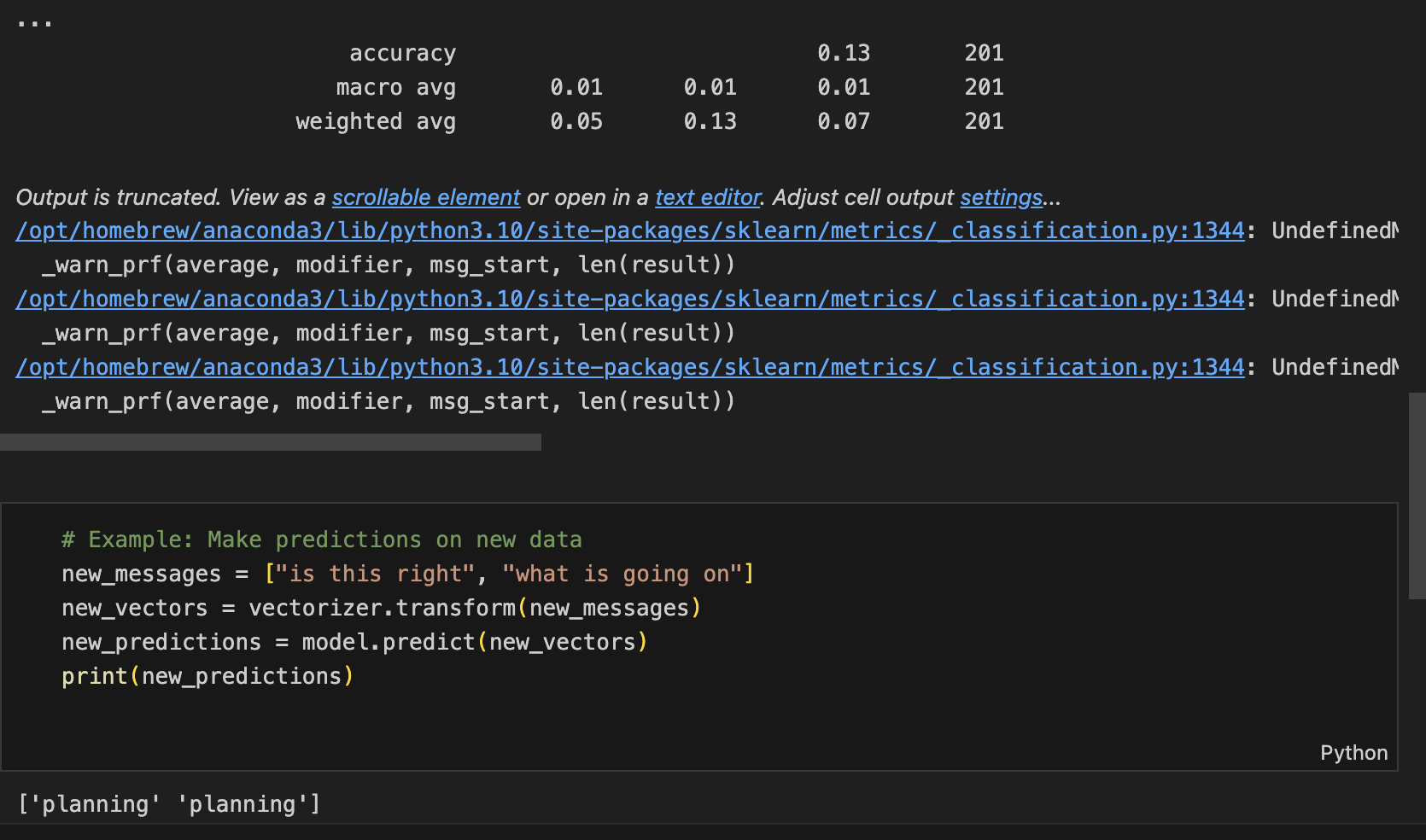


Fig 1. Predictive model Results

Hence, we then extract verbs from a dataset, analyze their frequency using n-grams, and perform clustering. We'll use the NLTK library for tokenization and n-gram generation, as well as scikit-learn for clustering.

We wrote a script that loads the dataset, tokenizes each text, extracts verbs using part-of-speech tagging, generates n-grams of verbs, counts the frequency of each verb n-gram, and performs K-means clustering on the verb n-grams.

Clustering helps in identifying patterns and relationships between verb n-grams. Verb n-grams within the same cluster tend to occur together frequently in the dataset. Clusters can provide insights into common actions, behaviors, or topics represented by the verb n-grams.

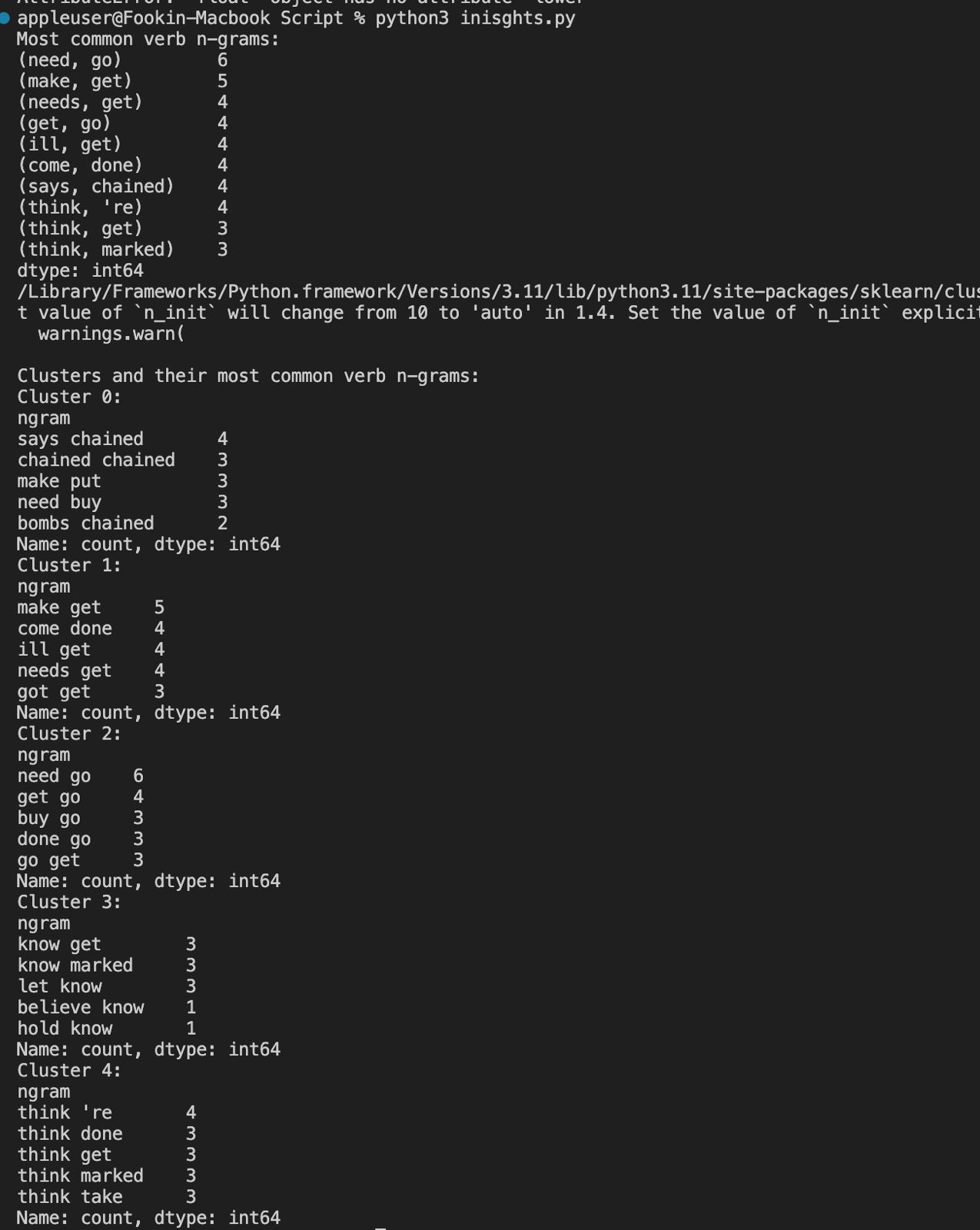
**Meaning and Interpretation:**

The output helps in understanding which verb combinations are most prevalent and how they can be grouped into meaningful clusters.

For example, if you have verb n-grams related to actions like "going," "buying," "trying," etc., these might indicate common activities or behaviors described in your dataset.

By analyzing these verb n-grams and clusters, you can gain insights into the underlying structure and content of your dataset, which can be useful for tasks such as topic modeling, behavior analysis, or content understanding.

1. **DOLL AI- Advisor Dataset Results**



**What do the results mean?**

The results indicate patterns in verb usage within the team communication dataset. Specifically:

Most Common Verb N-Grams: The most frequent verb combinations provide insights into the actions or tasks frequently discussed or executed by the team. For instance:

* (need, go): Indicates a recurring need or requirement to go somewhere.
* (make, get): Implies a common action of making or obtaining something.
* (says, chained): Suggests a communication style where statements are linked or connected.

Clusters and Their Verb N-Grams: Clustering reveals groups of verb combinations that tend to co-occur together, potentially representing specific planning strategies or communication patterns within the team. For example:

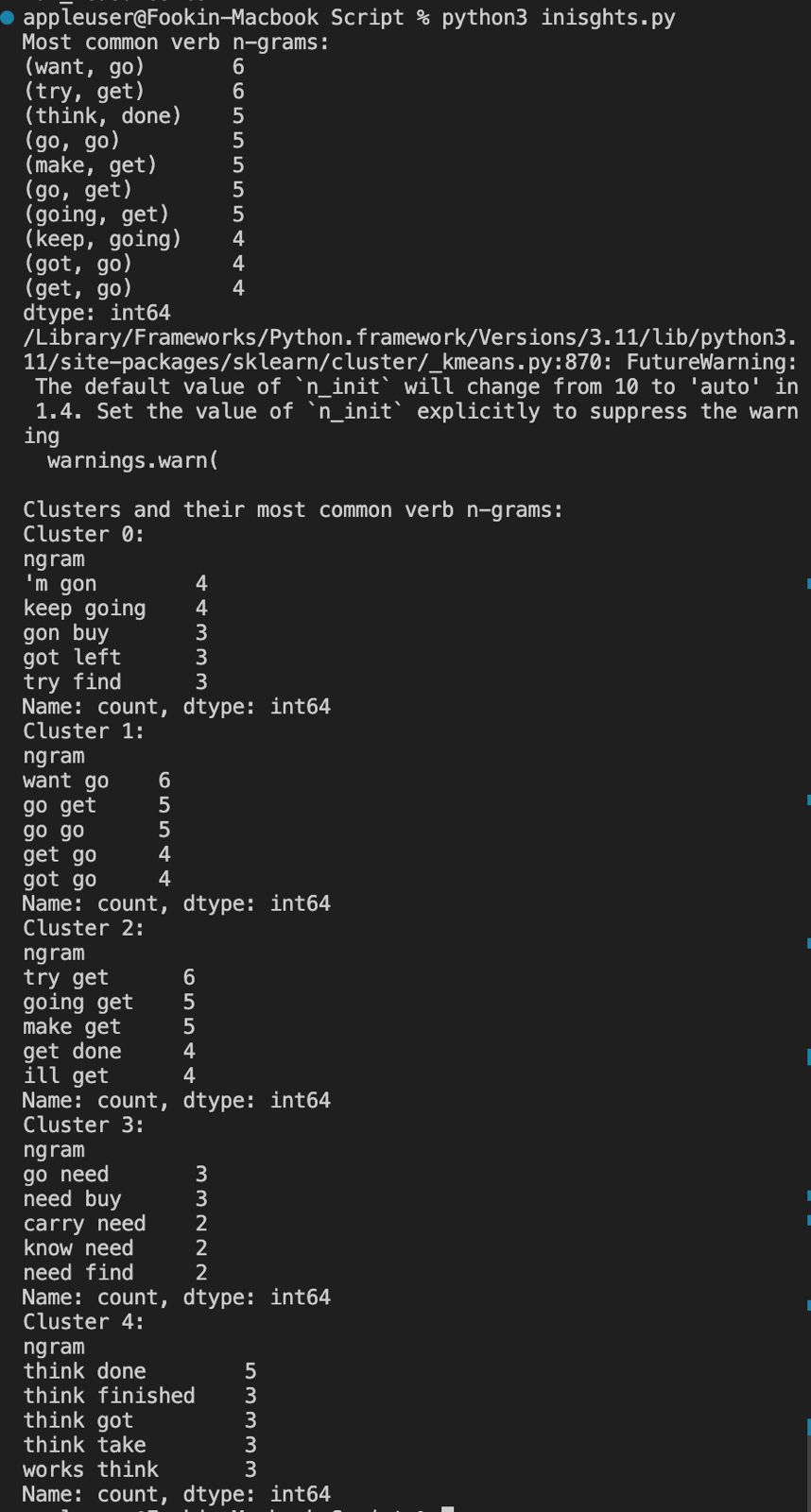
* Cluster 0: Contains verb n-grams like `says chained`, `chained chained`, `make put`, `need buy`, `bombs chained`, which might represent discussions about tasks, orders, or directives.
* Cluster 1: Shows verb n-grams related to various actions (`make get`, `come done`, `ill get`, `needs get`, `got get`), indicating a diverse range of planned activities or objectives.
* Cluster 2: Focuses on actions related to needs and actions (`need go`, `get go`, `buy go`, `done go`, `go get`), reflecting specific planning requirements or goals.

**What does it tell you about teams' planning strategies?**

The findings suggest several key insights into the teams' planning strategies:

* Action-Oriented Communication: The frequent verb n-grams highlight a focus on action-oriented discussions, emphasizing tasks, needs, and actions required to achieve goals.
* Diverse Planning Approaches: Clusters with varied verb combinations (e.g., Cluster 1) suggest diverse planning approaches, possibly reflecting different roles or responsibilities within the team.
* Communication Styles: Certain verb n-grams (e.g., `says chained` in Cluster 0) indicate specific communication styles or patterns, such as relaying instructions or linking information.

1. **CMU AI- Advisor Dataset Results**

****

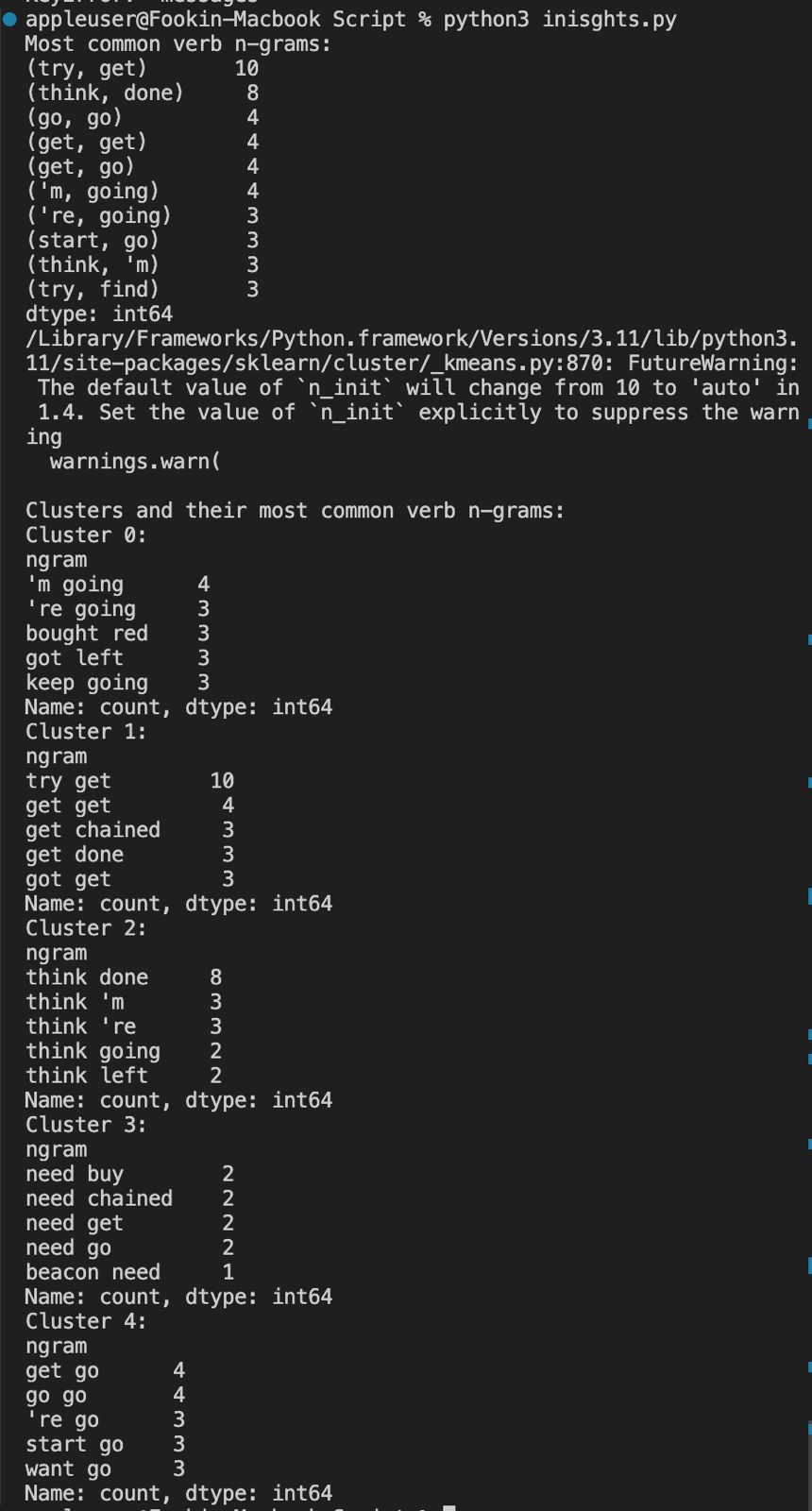
Interpreting these verb n-grams within the context of the team's mission or objectives is critical:

* Task Prioritization: Understanding which actions (`want go`, `try get`) are most frequently discussed sheds light on priority areas within the team's workflow.

Indicates desires or intentions aligned with actions related to movement or progress, suggesting a proactive approach towards achieving goals.

* Execution Strategies: Clusters (`Cluster 1`, `Cluster 2`) with diverse verb combinations suggest adaptive planning and execution strategies tailored to project requirements.
* Communication Patterns: Analyzing how verb n-grams reflect communication styles (`think done`, `keep going`) provides insights into effective collaboration and coordination practices.

1. **No Advisor Dataset Results**

****

The most common verb n-grams observed in this dataset include:

* (try, get): Indicates repeated attempts or efforts to achieve specific outcomes, highlighting a persistent approach to task completion.
* (think, done): Reflects cognitive processes leading to completion or decision-making, emphasizing thoughtful planning or problem-solving.
* (go, go), (get, get), (get, go): Signifies actions related to movement, progress, and task execution, demonstrating active engagement in project activities.
* Cluster 0 predominantly focuses on ongoing actions like `'m going`, `'re going`, and discussions related to specific tasks (`bought red`, `got left`), suggesting continuous progress and effective task management within the team.
* Cluster 1 is characterized by a mix of persistent efforts (`try get`) and task-oriented activities (`get done`, `got get`), highlighting a proactive approach to achieving objectives.
* Cluster 2 emphasizes cognitive processes (`think done`, `think 'm`, `think 're`) and decision-making (`think going`, `think left`), underscoring deliberate planning and strategic thinking.
* Cluster 3 centers around verbs indicating urgency (`need buy`, `need chained`, `need get`, `need go`) and resource allocation essential for task completion.
* Finally, Cluster 4 showcases actions (`get go`, `go go`, `'re go`, `start go`, `want go`) linked to task execution and progress, reflecting active engagement and goal-oriented behaviors within the team's communication dynamics.

These clusters collectively offer valuable insights into the nuanced planning approaches and coordination strategies adopted by human teams in their collaborative endeavors.

By analyzing verb usage and clustering findings, organizations can leverage these insights to improve team performance, streamline processes, and achieve mission success in diverse operational contexts.

**Discussion**

**Data Issues and Limitations**

There are many challenges in this project.

First is data cleaning. When participants say things, only some sentences are grammatically correct. In addition, the dataset automatically broke one sentence into several parts, which creates difficulty in coding and counting.

Second, some of the themes are not clear-cut. For example, coordination could overlap with commands. Seeking clarification may overlap with inquiry; however, seeking clarification is a specific type of inquiry.

Third is the analysis approach. We originally started with the manual coding approach. In this approach, two coders had extremely low agreement rates, and we did not proceed with this method due to new instructions to explore more advanced data analysis methods. In addition, there is a lack of a team planning framework to categorize the themes we came up with from the dataset. In the AntConc approach, we could not find a solution to find all the verbs in the dataset. The internet’s solution and using verbs.csv did not help. In the machine learning approach, though the chatGPT can automatically process a large dataset, the current chatGPTmodel has low accuracy in classifying the chat text.

**Insights**

Natural Language Processing at scale is still challenging.

Using the most frequently used verbs could indicate the capability of the team.

Verb frequency hints at team collaboration in bomb disposal.

Most used verbs reveal communication and task sharing.

The most frequently used words were still less than ⅓ of the trials in an advisor condition, which means ⅔ of the data may not have team communication. Therefore, the dataset quality is not so good.

**Future Directions**

The codebook developed in manual coding could be improved and used to find some way to improve classification accuracy and inter-rater reliability.

Explore more advanced machine learning algorithms to train the models, such as tensor flow (i.e., pick a subset of data to train the model, and then use the rest to test the model).

**Appendices**

**Appendix A: Stop List**

a

about

above

after

again

against

all

am

an

and

any

are

aren't

as

at

be

because

been

before

being

below

between

both

but

by

can't

cannot

could

couldn't

did

didn't

do

does

doesn't

doing

don't

down

during

each

few

for

from

further

had

hadn't

has

hasn't

have

haven't

having

he

he'd

he'll

he's

her

here

here's

hers

herself

him

himself

his

how

how's

i

i'd

i'll

i'm

i've

if

in

into

is

isn't

it

it's

its

itself

let's

me

more

most

mustn't

my

myself

no

nor

not

of

off

on

once

only

or

other

ought

our

ours

ourselves

out

over

own

same

shan't

she

she'd

she'll

she's

should

shouldn't

so

some

such

than

that

that's

the

their

theirs

them

themselves

then

there

there's

these

they

they'd

they'll

they're

they've

this

those

through

to

too

under

until

up

very

was

wasn't

we

we'd

we'll

we're

we've

were

weren't

what

what's

when

when's

where

where's

which

while

who

who's

whom

why

why's

with

won't

would

wouldn't

you

you'd

you'll

you're

you've

your

yours

yourself

yourselves

**Related References**

Ambrose, S. C., Matthews, L. M., & Rutherford, B. N. (2018). Cross-functional teams and social identity theory: A study of sales and operations planning (S&OP). *Journal of Business Research*, *92*, 270–278.

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Arbanas, B., Ivanovic, A., Car, M., Orsag, M., Petrovic, T., & Bogdan, S. (2018). Decentralized planning and control for UAV–UGV cooperative teams. *Autonomous Robots*, *42*(8), 1601–1618.<https://doi.org/10.1007/s10514-018-9712-y>

Brafman, R. I., & Domshlak, C. (2013). On the complexity of planning for agent teams and its implications for single agent planning. *Artificial Intelligence*, *198*, 52–71.

DeChurch, L. A., & Haas, C. D. (2008). Examining Team Planning Through an Episodic Lens: Effects of Deliberate, Contingency, and Reactive Planning on Team Effectiveness. *Small Group Research*, *39*(5), 542–568.<https://doi.org/10.1177/1046496408320048>

Evain, J. ‐N., Perrot, A., Vincent, A., Cejka, J. ‐C., Bauer, C., Duclos, A., Rimmelé, T., Lehot, J. ‐J., & Lilot, M. (2019). Team planning discussion and clinical performance: A prospective, randomised, controlled simulation trial. *Anaesthesia*, *74*(4), 488–496.<https://doi.org/10.1111/anae.14602>

Fischer, J. E., Greenhalgh, C., Jiang, W., Ramchurn, S. D., Wu, F., & Rodden, T. (2021). In‐the‐loop or on‐the‐loop? Interactional arrangements to support team coordination with a planning agent. *Concurrency and Computation: Practice and Experience*, *33*(8), e4082.<https://doi.org/10.1002/cpe.4082>

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied Psychology*, *99*(3), 423.

Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G., & Vathis, N. (2015). Heuristics for the time dependent team orienteering problem: Application to tourist route planning. *Computers & Operations Research*, *62*, 36–50.

Hakli, R. (2017). Cooperative Human–Robot Planning with Team Reasoning. *International Journal of Social Robotics*, *9*(5), 643–658.<https://doi.org/10.1007/s12369-016-0377-4>

Huang, L., Fouse, A., Cooke, N., & Weiss, E. (2024). *Artificial Social Intelligence for Successful Teams (ASIST) Study 4 Dragon Testbed Dataset* (Version 2) [dataset]. ASU Library Research Data Repository.<https://doi.org/10.48349/ASU/ZO6XVR>

Jeffs, L., Merkley, J., Sinno, M., Thomson, N., Peladeau, N., & Richardson, S. (2019). Engaging stakeholders to co-design an academic practice strategic plan in an integrated health system: The key roles of the nurse executive and planning team. *Nursing Administration Quarterly*, *43*(2), 186–192.

Kaw, N., Murray, J., Lopez, A. J., & Mamdani, M. M. (2020). Nursing resource team capacity planning using forecasting and optimization methods: A case study. *Journal of Nursing Management*, *28*(2), 229–238.<https://doi.org/10.1111/jonm.12905>

Kim, J. (2018). Intelligent Decision Support for Human Team Planning. *IJCAI*, 5769–5770.<https://www.ijcai.org/proceedings/2018/0828.pdf>

Kim, J., & Shah, J. A. (2017). Towards intelligent decision support in human team planning. *2017 AAAI Fall Symposium Series*.<https://cdn.aaai.org/ocs/16019/16019-69894-1-PB.pdf>

Kim, J., Woicik, M. E., Gombolay, M. C., Son, S.-H., & Shah, J. A. (2018). *Learning to infer final plans in human team planning*.<https://dspace.mit.edu/handle/1721.1/125887>

Kinnerk, P., Kearney, P. E., Harvey, S., & Lyons, M. (2023). An investigation of high-performance team sport coaches’ planning practices. *Sports Coaching Review*, *12*(3), 253–280.<https://doi.org/10.1080/21640629.2021.1990653>

Kollmann, T., Stöckmann, C., & Linstaedt, J. W. (2019). Task Conflict, Narcissism and Entrepreneurial Capability in Teams Planning a Business: A Moderated Moderation Approach to Explaining Business Planning Performance. *Journal of Small Business Management*, *57*(4), 1399–1423.<https://doi.org/10.1111/jsbm.12418>

Kruse, S. D. (2001). Creating communities of reform: Continuous improvement planning teams. *Journal of Educational Administration*, *39*(4), 359–383.

*Laurence Anthony’s Software*. (n.d.). Retrieved April 26, 2024, from<https://laurenceanthony.net/software.html>

Lehmann-Willenbrock, N., Meyers, R. A., Kauffeld, S., Neininger, A., & Henschel, A. (2011). Verbal Interaction Sequences and Group Mood: Exploring the Role of Team Planning Communication. *Small Group Research*, *42*(6), 639–668.<https://doi.org/10.1177/1046496411398397>

Lubejko, B. G. (2014). The Planning Team—Who Belongs? *The Journal of Continuing Education in Nursing*, *45*(6), 244–245.<https://doi.org/10.3928/00220124-20140527-11>

Mehta, A., Feild, H., Armenakis, A., & Mehta, N. (2009). Team Goal Orientation and Team Performance: The Mediating Role of Team Planning. *Journal of Management*, *35*(4), 1026–1046.<https://doi.org/10.1177/0149206308326773>

Nagy, G. A., Cassiello-Robbins, C., Anand, D., Arnold, M. L., Coleman, J. N., Nwosu, J., Singh, R. S., & Woodward, E. N. (2022). Building a multicultural peer-consultation team: Planning, implementing, and early sustainment evaluation. *Transcultural Psychiatry*, *59*(6), 844–862.<https://doi.org/10.1177/13634615221105117>

Oldeweme, M., Konradt, U., & Garbers, Y. (2021). Effects of Situational Factors on Team Planning: A Policy-Capturing Approach. *Zeitschrift Für Arbeits- Und Organisationspsychologie A&O*, *65*(2), 68–81.<https://doi.org/10.1026/0932-4089/a000348>

Romyn, D., & Kebbell, M. (2014). Terrorists’ planning of attacks: A simulated ‘red-team’ investigation into decision-making. *Psychology, Crime & Law*, *20*(5), 480–496.<https://doi.org/10.1080/1068316X.2013.793767>

Sibbald, S. L., Van Asseldonk, R., Cao, P. L., & Law, B. (2021). Lessons learned from inadequate implementation planning of team-based chronic disease management: Implementation evaluation. *BMC Health Services Research*, *21*(1), 134.<https://doi.org/10.1186/s12913-021-06100-4>

Sreedharan, S., Kulkarni, A., & Kambhampati, S. (2022). Explainable Human-AI Interaction: A Planning Perspective Sreedharan. In *Synthesis Lectures on Artificial Intelligence and Machine Learning: Lecture# 47*. Morgan and Claypool Publishers.<https://asu.elsevierpure.com/en/publications/explainable-human-ai-interaction-a-planning-perspective-sreedhara>

Szwarc, E., Nielsen, I., Smutnicki, C., Bocewicz, G., Banaszak, Z., & Bilski, J. (2021). Competence-oriented project team planning – university case study. *Journal of Information and Telecommunication*, *5*(3), 310–333.<https://doi.org/10.1080/24751839.2020.1857039>

Szwarc, E., & Wikarek, J. (2020). Proactive Planning of Project Team Members’ Competences. *Foundations of Management*, *12*(1), 71–84.<https://doi.org/10.2478/fman-2020-0006>

Traynor, K. (2017). Planning, teamwork made Harvey manageable for Houston hospitals. *American Journal of Health-System Pharmacy*, *74*(21).<https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10792082&asa=N&AN=125885785&h=TOihnI%2FC%2BNATKhkgf%2FIXuBBJxydnr%2BuKDiZJ0lSwM58WyYdq8OeJeWq0c%2BghJHTPaOyri1XPV04h4j197Wd4Sw%3D%3D&crl=c>

Ure, N. K., Chowdhary, G., Chen, Y. F., How, J. P., & Vian, J. (2014). Distributed Learning for Planning Under Uncertainty Problems with Heterogeneous Teams: Scaling Up the Multiagent Planning with Distributed Learning and Approximate Representations. *Journal of Intelligent & Robotic Systems*, *74*(1–2), 529–544.<https://doi.org/10.1007/s10846-013-9980-x>

Walker, J. S., & Schutte, K. (2005). Quality and Individualization in Wraparound Team Planning. *Journal of Child and Family Studies*, *14*(2), 251–267.<https://doi.org/10.1007/s10826-005-5052-6>

Walker, J. S., Seibel, C. L., & Jackson, S. (2017). Increasing Youths’ Participation in Team-Based Treatment Planning: The Achieve My Plan Enhancement for Wraparound. *Journal of Child and Family Studies*, *26*(8), 2090–2100.<https://doi.org/10.1007/s10826-017-0738-0>

Wallace, H. (2021). Planning in professional learning teams: Building trust, common language and deeper understanding of pedagogy. *The Australian Educational Researcher*, *48*(2), 377–395.<https://doi.org/10.1007/s13384-020-00394-9>

Willox, S., Morin, J., & Avila, S. (2023). Benefits of individual preparation for team success: Planning for virtual team communication, conflict resolution and belonging. *Team Performance Management: An International Journal*, *29*(1/2), 1–14.

Woolley, A. W., Gerbasi, M. E., Chabris, C. F., Kosslyn, S. M., & Hackman, J. R. (2008). Bringing in the Experts: How Team Composition and Collaborative Planning Jointly Shape Analytic Effectiveness. *Small Group Research*, *39*(3), 352–371.<https://doi.org/10.1177/1046496408317792>

Zhao, Y. (Audrey), Srite, M., Kim, S., & Lee, J. (2021). Effect of team cohesion on flow: An empirical study of team‐based gamification for enterprise resource planning systems in online classes. *Decision Sciences Journal of Innovative Education*, *19*(3), 173–184.<https://doi.org/10.1111/dsji.12240>