Live Analysis of Security Operations Center (SOC) Analysts in Training

Michael Tsiderdekis, Jacob Robinson, Jacob McCabe, Rachid Soro, Amandeep Kaur

***Abstract*—The analysis focuses on the ticket data from undergraduate students of security operations centers (SOC) from various universities. The aim of this analysis is to improve both the educational process and the security of local municipality networks. A group of four students participated in this research. During the live ticket analysis, the team focused on addressing two main factors: those that played a role to increase the ticket participation and those reducing the ticket resolution time. Results are achieved by examining the real time data and ticket handling behavior of the SOC Analysts trainees. This research on the ticket management system has the potential to improve both the ticket participation and resolution time. Overall, the research on ticket management systems in the context of SOC can greatly improve the ticket-handling capabilities of organizations in different sectors and industries. The team found that while there may not be a relationship between ticket fields and the resolution time, there is the potential for determining the resolution time from aspects such as the description or number of ticket actions. Analysis of ticket participation shows meaningful relationships between a user's login count and their participation in the ticket management system. This knowledge can be used to create better practices and guidelines for SOC Analysts, improve risk mitigation, and to enhance network security.**

***Index Terms*—Security operations center (SOC), intrusion detection system (IDS), ticket, ticket resolution, ticket participation**

I. Introduction

In most organizations there exists some form of ticket and ticket resolution system [1]. The system may exist to serve customer service, logistics, or even research and development applications. A successful outcome to a ticket requires both ticket participation, and fast ticket resolution. Additionally, ticking systems are also ideal for creating, monitoring, and escalating network security events with collaborative investigations and supporting evidence. The research described in this paper is based off such a scenario.

More specifically the data is based off undergraduate student driven security operations center, or SOC, ticket data from multiple universities and class years [1], [2], [3], [4]. While the data is primarily based on student activity, or trainee analysts, the network traffic logged by Intrusion Detection Systems (IDS) are authentic and aim to support multiple small municipality networks across the United States. Therefore, any findings and future research will not only help improve the academic process, but moreover, improve the security of these networks [2].

II. Related Works

A. Human Factor of SOC

In the context of cybersecurity research, the technological factor has been the focus of many studies. Due to this, the human aspect has less research surrounding it [3]. Regardless of the technical specifications of an IDS, it requires a person to investigate an event and handle ticket management.

One way to booster the effectiveness of SOCs is to improve the evaluation metrics of analysts. According to Agyepong et al, there is a lack of metrics for evaluating SOC analysts [5]. Most current metrics focus on quantitative values rather than qualitative values. Chamkar et al claim that the lack of qualitative feedback results in dissatisfaction from analysts [4]. These works suggest that it is important to determine what it is that makes a successful ticket. Such a basis would provide improvements to SOC analyst education and evaluation.

There are several significant roadblocks identified via a survey of SOC analysts by Chamkar et al. The top roadblocks identified include a lack of automation and orchestration of small security tasks, which would have the capability of freeing up crucial time for analysts to focus on more nuanced tasks.

Another issue identified by the survey included a large number of false positives. Since tickets must be handled by an analyst, similar to a lack of automation, handling false positives create excess work for analysts that can leave the system vulnerable to real attacks.

B. Collaborative Learning

Collaboration is seen in all aspects of life: education, work, socialization, and more. The promotion of collaborative learning is not always seen as critical to education, but it is seen as necessary for developing many skills in higher education [6], [7].

This form of learning has two important aspects, student-to-student and student-staff; both are important roles to confer in a classroom for both the students' education and the teachers' pedagogic practices [7]. According to Frykedal and Chiriac, these processes promote active participation in inclusive, analytical discussions when supplemented by the teacher as a contributing member [8].

Assessment of collaborative learning environments lead to roadblocks established in research. A study by Le et al shows that focusing on cognitive aspects of collaborative learning can cause teachers to ignore that collaborative aspects as a result of goal setting and student assessment [9]. While collaborative learning has the potential for improving students' ability to learn by promoting inclusion, there are hindrances that may stand in the way.

III. Research Design

A. Research Questions

The two research questions driving this study include:

1) What leads to a faster resolution of a ticket?
2) What leads to increased ticket participation on a ticket?

B. Data Analysis

The data set driving this research is confidential, and is unavailable for the public due to its sensitivity. Additionally, there is little previous research completed on the data set. Previous research has, however, shown that over the course of a given class resolution time of a ticket has decreased.

The bulk of the data set can be described as ticket meta data. And while it does contain student information it is often limited, or not filed consistently. For example only about two thirds of all participating students have a valid ”edu” email address. This makes categorizing students based on University and time of attendance extremely challenging. To combat this challenge any categorization of data has had to be done creatively, but also with great care.

In the raw data set, there are 3532 tickets with 4941 total actions spread across 785 users. This represents the total values observed in the data set over a four year and eight month time period with nine participating academic institutions. It should be noted that these values are prior to any data clean up. The composition of the 785 users is known to be administrative users, professors and instructors, and students. It is unknown which users belong to which category. During investigations into specific hypothesis the data was restricted to eliminate noise and non-students as much as possible.

While not mentioned in full detail in this report, early in the research phase an attempt was made to look at student behavior in blocks of time based on ticket activity. If ticket creation was halted for a period of time, especially if it was near the time of a traditional academic seasonal transition, a cut off was made. When looking at behavior in these sections of time similar trends of behavior were observed, for example number of tickets made per class(es), were relatively consistent. But at times it was difficult to make a cut off date since there exists a discrepancy of class duration, size, or start/end date between participating schools. This finding mirrored previous work on this data set while it was smaller.

However, most of other aspects of the data set were far from being normally distributed which is to be expected with data based solely on human behavior. Moreover, because of the size of the data set and notable similarities between time periods, any observations or findings were relevant to making actionable changes to the academic process.

C. Data Categorization

*1) **Ticket resolution time**:* The amount of time it takes for a ticket to resolve varies drastically. Some tickets were marked closed within a day and some tickets took over two years after creation to be resolved. In order to classify ticket resolution time, certain ticket fields were identified to be important features. These features included ”Summary Length”, ”Description Length”, ”Steps to Reproduce Length”, ”Additional Information Length”, and ”Number of Actions”. Due to the noise in the raw data, data points whose features with values in the 99^{th} percentile were considered outliers, and dropped from the data set. The remaining data points were then standardized to make future computations more reliable.

There will be two possible binary classifications of resolution time to explore. The first is ”within a school term (10 weeks)” and ”longer than a school term”. The second classification will be ”within an academic year (30 weeks)” and ”longer than an academic year”.

To decide on a classification, consider the following hypothesis: *there exists a significant difference in the mean between the resolution categories for each ticket feature*. Then the null hypothesis follows as: *there does not exist a significant difference in the mean between resolution categories for each ticket feature*. To determine which categorization to use, choose the one with more significant features at a 95% confidence level.

P-Values	Summary	Desc	Repro	Add’l Info	Actions
Term	5.66e-3	5.34e-4	0.838	0.0623	1.363-10
Year	1.12e-3	1.51e-2	0.435	1.55e-2	<2e-16

TABLE I

From the hypothesis test, classifying resolution time by the academic year results in four ticket features having significant p-values while classifying by the academic term only results in three ticket features having significant p-values (see Table I). As a result of this, the categorization of ticket resolution time will be defined into the binary categories: ”within the academic year” and ”longer than an academic year”.

*2) **Ticket Participation**:* Data analysis on the topic of participation went through two primary phases.

1) Ticket characteristics: Data analysis around ticket characteristics followed a similar formula to ticket resolution time. Initially any aspects of a ticket and meta data of a ticket were individually tested against the number of actions on a particular ticket.

Length of a Title	If “ET” is in the title
The Length of the Description	No. of external links in the Description
Does the additional info field exist	No. of external links in Additional Info
Does steps to reproduce exist	

TABLE II

While some of these attributes are obvious, it is worth discussing their contextual importance. (Table II). Emerging Threats (ET) is the act of a student stating what ET signature they have encountered in the title of the ticket. ET is the IDS rule set that is used by the students to quickly identify potential threats on the network they are responsible for. When a student views alerts they will be followed by an ET signature. ET in the title of a ticket represents the preferred behavior for tilting a ticket. The use of ET followed by the threat description is the easiest way to describe the context and suspicious activity a student witnesses. Tickets without this distinction can be harder to process quickly due to their inherent vagueness. Moreover, using the ET database, it is easy to view what behavior has triggered the threshold for the event, which not only gives context on that potential threat, but gives the student an idea what potential cyber threat behavior looks like. The description field of a ticket is the primary field for a student to describe important data points of the event they witness, such as time of event, amount of data related to the event, and any explanation of the validity or importance of a potential threat. The importance of external links in additional information or description fields shows what amount of investigation a student may have completed about a threat using open source intelligence or OSINT. OSINT being any information that can be gathered from publicly available sources [10]. While the world of OSINT is large and its applications even larger, for the context of the SoC student there is a diverse, but much narrower, landscape of OSINT platforms. Typically, students use any OSINT platform that allows the lookup of an external IP address, domain name, or port usage in relation to cyber threats. While these platforms do have their own range of information and purpose they all aid the student investigation process for determining what is likely a cyber threat or benign network traffic. Because of this, it was fair to assume that the more links present or lack thereof may impact how many ticket actions take place. For example, if no external links exist it leaves an opportunity for students to help investigate a threat, and comment accordingly. Additionally, because of the amount of tools that exist for a student it makes additional comments with their own research that is unique to prior research. Steps to reproduce is a field where a student is meant to post links related to viewing the potential threat they have discovered. These could range from a singular link or multiple, the number of links is less relevant to the OSINT related links since the steps to reproduce links are typically from one platform. Moreover, there could be other data types such as screen caps, so for this research the primary concern is just that the field exists and is not null. All of these tests used Spearman's rank correlation coefficient. For brevity, the only significant finding is that no single characteristic of ticket meta-data was correlated with the number of actions on a particular ticket. Following this there was an attempt of using a multiple logistic regression of multiple characteristics tested previously. However, some data clean up did take place before the test was conducted (see Table III).

No. of Unique Users	No. of Bug Actions
$n \geq 2$	$2 \geq n \leq 30$

TABLE III

These restrictions are valid for two primary reasons. Firstly, the limitation of having at least two unique users on a ticket helps remove tickets in the data set that are tests. Additionally, it removes the majority of tickets that were created near the end of a data set that were created just before the data was collected. These tickets appear to be made right after a class had started and little to no time passed. Secondly, while the majority of tickets have much less than 30 actions, there did exist outliers far beyond 30 and it felt necessary to remove them since they did not represent typical ticket behavior between classes.

Additionally, the effects of these characteristics would be tested against its effect on the number of unique users and number of actions on a ticket separately.

2) User behavior: User behavior is easier to define than the previous test. There exists one primary question, *what is the relationship between how many tickets a user creates and how many actions are done by a user?* This data point also required removal of noise (see Table IV). Primarily when determining when a user is a student or a instructor. Some users were already identified as such, however, because of the expansion of the data set more investigation would be required. Any users with little to no tickets, but an abnormally high amount of actions were eliminated since these users are likely to be a instructor commenting on tickets. Furthermore, any users with a email associated with the critical insight organization, or contained "instructor" were also removed. This along with the users noted by previous research, provided confidence that only student actions were isolated for consideration. Additionally, like the previous test, if a user had no activity, they were removed from consideration.

Tickets	Users	Actions
452	458	1798

TABLE IV

With data limited to just student activity, they can be categorized by students who just make tickets and users who make tickets and comments. These two categories represent nearly a 50:50 ratio. These two student behavioral outcomes were then tested against a students login count to see if there is a correlation between user login and the outcome of their activity. Furthermore, this distinction of behavior allows important metrics to be declared.

Avg. Ticket Actions	Avg. Tickets
11	7

TABLE V

It should be noted that Table V represent the average per student, and the average Ticket Actions includes ticket creation as a action. Because the research question is focused on higher participation, it felt as determining what may lead to users who participate with actions and tickets is critical.

IV. Results

A. *How do ticket characteristics affect resolution time (RQ1)?*

Before creating the multiple logistic regression model, the data cleaned for resolution time was split into train and test sets using a randomized 70%-30% split. This was done in order to prevent data leakage, allowing for a better gauge of accuracy. The first model considered is simply using all of the features of a ticket identified by the team during the categorization of ticket resolution time. The second model will be creating a nonlinear term from the features ”Steps to Reproduce Length” and ”Ticket Actions”. Training both models gave very similar results; the AIC values were 2327.1 and 2325.9, respectively. As a result, only consider the second model since it was the smaller of the two AICs.

To determine how well the model performs, it is necessary to create a baseline for comparison. The baseline test is defined by predicting the majority class from the training set on the test set. In this case, it was found that the majority class in the training set is the tickets resolved within an academic year. Predicting that every ticket in the test set belongs in this category results in an accuracy of 77.61%.

	Predicted False	Predicted True
Actual False	29	142
Actual True	46	547

TABLE VI

The results of using the model to make predictions on the test set are seen in Table VI. The accuracy of the final model is 75.39% and the F1 score is 85.33%. Comparing its accuracy with the baseline shows that it is a poor model since the baseline performed better. This means that the multiple logistic regression was unable to produce a relationship between the given features and the time it takes to resolve a ticket.

While the model did not perform well, previous testing shows the possibility for the features to hold predictive power over ticket resolution time. Perhaps with added features describing the textual contents of the ticket fields or more data could produce better results.

B. *What behavior leads to faster resolution time (RQ1)?*

To address this research question, a data set of tickets was analyzed, focusing on resolved tickets exclusively. By filtering the data set, two categories were established: tickets resolved ”faster” (within 7 days) and tickets resolved ”slower” (beyond 7 days). Word count in ticket summaries and the involvement of different analysts were examined as potential factors influencing resolution time. Descriptive statistics and comparative analysis were employed to explore (please refer to Table VII) any associations between these variables and resolution time. The analysis was conducted on a comprehensive data set consisting of 2,656 resolved tickets. These tickets were further classified into two distinct categories: ”Solved faster” and ”Solved slower.” Among the resolved tickets, 799 fell into the ”Solved faster” category, while a larger proportion of 1,857 tickets were categorized as ”Solved slower” (please refer to the Table VII for a detailed breakdown). Upon examining the word count in the ticket summaries, an interesting trend emerged. It was observed that the median word count in the ”Solved faster” category was noticeably higher compared to the ”Solved slower” category. This finding indicates that tickets with lengthier summaries, encompassing additional details and information, were more likely to be addressed and resolved a bit more faster.

Tickets solved faster	Tickets solved slower
799	1857

TABLE VII

This correlation between word count and resolution speed suggests that when tickets contain more comprehensive information, it facilitates expedited resolution. The additional details provided in the ticket summaries may enable the analysts to quickly understand the issue at hand and initiate appropriate actions, resulting in faster resolutions. Moreover, a higher word count could indicate that the customers have provided more context or background, which aids in problem diagnosis and resolution. In contrast, the analysis of character count alone did not uncover any substantial disparities between the two categories. The median character count remained relatively similar in both the ”Solved faster” and ”Solved slower” categories. Consequently, it can be inferred that the mere number of characters present in the ticket summaries does not play a decisive role in determining the overall resolution time.

However, it is important to note that character count alone might not capture the full complexity or level of detail in the ticket summaries. While two tickets may have similar character counts, one ticket might contain concise and precise information, while the other ticket may have a more detailed description. Therefore, focusing solely on character count may overlook valuable nuances that can impact resolution time. Delving deeper into the investigation, an exploration of the involvement of different analysts in the resolution process revealed a notable pattern. Specifically, one particular analyst stood out, identified by the unique Handler ID 88. This analyst proved to be instrumental in resolving a significant portion of the tickets across both the ”Solved faster” and ”Solved slower” categories. (please refer to Table VIII below for a detailed breakdown)

Handler ID	Ticket solved faster	Ticket solved slower
88	275	683
245	71	148
399	54	40
91	48	112

TABLE VIII

The exceptional efficiency demonstrated by this analyst in ticket resolution could potentially be attributed to their extensive experience in the field. Having encountered a wide range of issues and scenarios over time, this analyst most likely possesses deep domain knowledge and a strong understanding of the systems and processes involved. As a result, they may possess a heightened ability to quickly identify solutions and execute them more effectively.

However, it is worth noting that this same analyst also handled the highest number of tickets falling into the "Solved slower" category. This observation highlights the fact that although experience can undoubtedly contribute to faster resolutions, it is not the sole determining factor in guaranteeing timely resolutions for all cases. To have better view of how much the analysts are contributing, the following table will show the percentage of how many tickets of each category, each analyst has solved. (please refer to Table IX below for a detailed breakdown)

Handler ID	Solved Faster(Percentage)	Solved Slower(Percentage)
88	34.42%	36.78%
245	8.89%	7.97%
399	6.76%	2.15%
91	6.00%	6.03%

TABLE IX

Other factors, such as the complexity of the issue or the availability of required resources, may also impact resolution time. It is possible that the tickets assigned to Handler ID 88 in the "Solved slower" category involved more intricate problems that necessitated additional time for investigation or collaboration with other teams. Consequently, even with extensive experience, there may be limitations to the analyst's ability to consistently achieve faster resolutions.

In summary, the comprehensive analysis of the data set provides valuable insights. It suggests that tickets with higher word counts in their summaries tend to be resolved at a faster pace, implying that the inclusion of detailed information aids in expediting the resolution process. The additional context and specifics provided in these tickets enable the analysts to swiftly comprehend and address the underlying issues.

Additionally, the participation of experienced analysts, as exemplified by the analyst with Handler ID 88, appears to play a significant role in achieving faster resolutions. Their extensive knowledge and expertise likely contribute to their ability to quickly diagnose and resolve a wide range of issues. However, it is important to acknowledge that resolution time can still vary based on the complexity and nature of the problem.

To gain a comprehensive understanding of the relationship between different factors and resolution times, further investigation and analysis are necessary. Exploring additional variables, such as the experience or knowledge between analysts, the recurrence of certain specific words and the collaboration between analysts and other teams, may provide a more nuanced view of the factors influencing resolution time. By identifying these underlying dynamics, it could be much easier to support processes and enhance the overall efficiency/speed of ticket resolution.

C. *What behaviors hinder ticket resolution (RQ1)?*

Ticket resolution rate highly depends on the factors such as providing enough information within the ticket, routing the ticket to the appropriate location, resolving, and closing the tickets correctly. Research demonstrates that some of the tickets were sent to the wrong location and some of the tickets did not have enough information to proceed with the ticket, which means that there could be a potential issue with the current guidelines for submitting tickets, or that the ticket support team may need to offer additional resources such as templates or forms to assist reporters in providing more comprehensive and precise information when submitting tickets. And there may also be underlying problems with the ticket classification or routing system. This could result in prolonged resolution times and concerns regarding the tasks not getting completed in a timely manner due to incorrect routing.

In addition, some of the tickets remained unresolved by the ticket handlers because they were unable to effectively close these tickets. Further investigation into this matter showed that the handlers were not equipped with the required skills, training, and right access privileges to successfully finish the resolution process for these specific tickets.

To determine what percentage of these situations affected the overall ticket resolution time. The entire data set is analyzed based upon the percentage of keyword occurrences. Data was gathered related to the keyword categories: Marked Resolved, Wrong Location, Lack of Information, Unable to Close and Closed Tickets. Of the total number of tickets in the data set, 39% were marked as closed and 13% were marked as resolved. On the flipped side, it was determined that 35% of total tickets were lacking important information, 10% of tickets were sent to the wrong location, and 3% of tickets were unable to be closed. Some of the tickets that remain open could possibly be new tickets that were not handled at the time of analysis.

D. *What affect does ticket characteristics have on greater participation (RQ2)?*

Outcomes from preliminary testing and data analysis pointed that perhaps a multiple linear regression include all variables listed in the design section would in fact show a correlation between specific ticket characteristics and greater participation. Or more broadly, that all of the variables together will show to have an effect on the outcome of a ticket. However, both for number of actions on a ticket and for the number of unique users commenting on a ticket there was no statically relevant correlation.

	No. of Unique Users	No. of Ticket Actions
Adjusted R^2	0.061	0.039
p-value	2.2e-16	2.2e-16

TABLE X

Based on the adjusted R^2 values (See Table X) for both the number number of unique users and number of bug actions, the model presented using the previously mentioned ticket characteristics do not act as a model to predict participation. However, it should be noted that as variables were added, the adjusted R^2 always increased. This supports the claim that the attributes tested do not serve as a reliable model for predicting better participation, but because of the consistent increase they do point to being independent variables that are correlated in someway. Potentially measuring these attributes with a different such as natural language processing on both tickets and comments can return better insights.

*E. **How does login count relate to greater participation (RQ2)?***

One problem when looking at behavioral trends with the data set is aspect of time. Every student maybe subject to different class lengths, and there is also the possibility of two students of the same length having vastly different productivity. To deal with this ratio was made for all users:

$$\frac{\text{Ticket Actions - Tickets Made}}{\text{Ticket Actions}}$$

This gives an average ratio of 0.27, or *on average 27% of users actions were not related to creating tickets*. Because of the limitations of the data set from the aspect of user statistics there is only one metric available that exists for all users which is *login count* (LC). Testing login count against the ratio of behavior, and the two behavioral categories.

	LC vs. Ratio	LC vs. All Actions	LC vs. Tickets
ρ	0.13	0.40	0.33
R^2	0.003	0.20	0.15
p-value	0.005	2.2e-16	7.5e-13

TABLE XI

There exists two takeaways from these findings shown in Table XI. Firstly, the login count vs. behavior ratio was expected to show some correlation, but the results show this not to be the case. Moreover, there does appear to exist a correlation between the other two behaviors. This does make sense given the amount of behaviors witnessed. Moreover, the fact that there is a correlation between login count and ticket creation as well as ticket actions does show that it's likely a student will contribute more if they login in more. Possibly the hardest part about the network analyst experience is the learning curve. Students often need to have a baseline knowledge of how networks work as well as the threat landscape, but the skills related to identifying, validating, and explaining real network threats requires practice and exposure with real network data. These findings support the idea of the more a student does the greater they can contribute given a class session. The less a student participates, measured by login count, the less the student will contribute to the learning experience and SoC analyst role. Additionally, there is another interpretation of these results is that it was expected that the Spearman correlation value to be extremely high, however this not the case. The coefficients are considered strong and moderate respectively. Moreover, the R^2 values are low, but given the context of irregular human data, the values are meaningful.

V. Conclusion

The live analysis of the ticket system provided valuable insight into both the ticket participation and ticket resolution that can be helpful to make the current overall ticket system better. The outcomes have capability to optimize the ticket management system, improve the current ticket handling procedures, mitigate the risks and strengthen the overall network security of the organizations.

Statistical testing on the data suggests that ticket fields are clustered based upon how long it takes for a ticket to be resolved, as seen in Table I. However with the current data, the team was unable to accurately predict how long it will take for a ticket to be resolved. There is potential for a better model if the textual contents of each ticket were to be flagged for keywords, resulting in a larger set of features. For example, these proposed features could include whether a ticket has the term "Emerging Threat" in the summary field or links referencing other sources in the description.

Moreover, in terms of attempting to bolster participation in terms of ticket actions, students appear to be graded primarily on a arbitrary sum of ticket creation and ticket actions, with both having equal value. Because of this and the finding that about half of all students only make tickets there is a possibility that changing the reward structure of the classes could promote better ticket participation. For example, a class could set a benchmark that a user is required to create at least two tickets and at least two actions/comments a week. This would ideally force students to spend time investigating existing tickets, and improving the students learning process while also resolving tickets quicker.

In terms of future work that can be conducted in the category of participation is how participation changes over a class. Given the unclear cut off dates of classes this will require more work on defining classes times. However, looking at the participation of users and on individual ticket over a class period may give insights on what is causing varying levels of participation.

References

[1] M. Majid and K. Ariffin, "Model for successful development and implementation of cyber security operations centre (soc)," *PLoS ONE*, vol. 16, no. 11, nov 2021.

[2] J. Marquardson, "Soc it to 'em: Bringing a security operations center onto a university campus," *Journal of Information Systems Education*, vol. 33, no. 3, pp. 300 –305, sept 2022.

[3] R. Dalal, D. Howard, R. Bennett, C. Posey, S. Zaccaro, and B. Brummel, "Organizational science and cybersecurity: abundant opportunities for research at the interface," *Journal of Business and Psychology*, vol. 37, pp. 1–29, 2022.

[4] S. Chamkar, Y. Maleh, and N. Gherabi, "The human factor capabilities in security operation center (soc)," *EDPACS*, vol. 66, no. 1, pp. 1–14, oct 2021.

[5] E. Agyepong, Y. Cherdantseva, P. Reinecke, and P. Burnap, "Challenges and performance metrics for security operations center analysts: a systematic review," *Journal of Cyber Security Technology*, vol. 4, no. 3, pp. 125–152, dec 2019.

[6] B. Liebech-Lien and E. Sjølie, "Teachers' conceptions and uses of student collaboration in the classroom," *Educational Research*, vol. 63, no. 2, pp. 221–228, nov 2020.

[7] C. Dickerson, J. Jarvis, and L. Stockwell, "Staff–student collaboration: student learning from working together to enhance educational practice in higher education," *Teaching in Higher Education*, vol. 21, no. 3, pp. 249–265, jan 2016.

[8] K. Frykedal and E. Chiriac, "Student collaboration in group work: Inclusion as participation," *International Journal of Disability, Development and Education*, vol. 65, no. 2, pp. 183–198, aug 2017.
[9] H. Le, J. Janssen, and T. Wubbels, "Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration," *Cambridge Journal of Education*, vol. 48, no. 1, pp. 103–122, jan 2017.
[10] R. Ghioni, M. Taddeo, and L. Floridi, "Open source intelligence and ai: a systematic review of the gelsi literature," *AI Society*, vol. 1-16, no. 1, pp. 1–2, jan 2023.