Developing Multi-Signal Phishing Detection and Continuous User Training Platform with Parsons Corporation

A Professional Readiness Experiential Program (PREP) Project Effort

----- ***Authors / Student Project Team Members*** -----

Taylor Le is a student at George Mason graduating with a bachelor's degree in information technology with a concentration in cybersecurity.

Tien Nguyen is a student at Virginia Tech graduating with a bachelor's degree in computer science.

Richard Nguyen is a student at Virginia Tech graduating with a bachelor's degree in computer science.

----- ***Industry Participant / Mentor*** -----

Nathan Dykas
Senior Computational Intelligence Engineer
Parsons Corporation

Daniel Boyce
AI/ML Engineer
Parsons Corporation

----- ***Faculty Member*** -----

Brian K. Ngac, PhD
FWI Corporate Partner Faculty Fellow
Assistant Dean, Centers of Excellence
George Mason University's Costello College of Business
bngac@gmu.edu

Interested in being an Industry Participant and or PREP Sponsor? Please reach out to bngac@gmu.edu, Thanks!

Introduction

Orion is a browser extension and desktop runtime system developed by a cross-institutional research team from George Mason University and Virginia Tech, in partnership with Parsons Corporation and the Commonwealth Cyber Initiative. The project addresses a gap that sits at the center of most phishing defense programs: the separation between detecting a threat and helping users understand it. Orion treats those two goals as one. It analyzes pages in real time as users browse, and when it detects phishing, it does not just raise an alarm. It walks users through exactly what made the page suspicious, tied to the specific elements on the screen in front of them.

Business Challenge

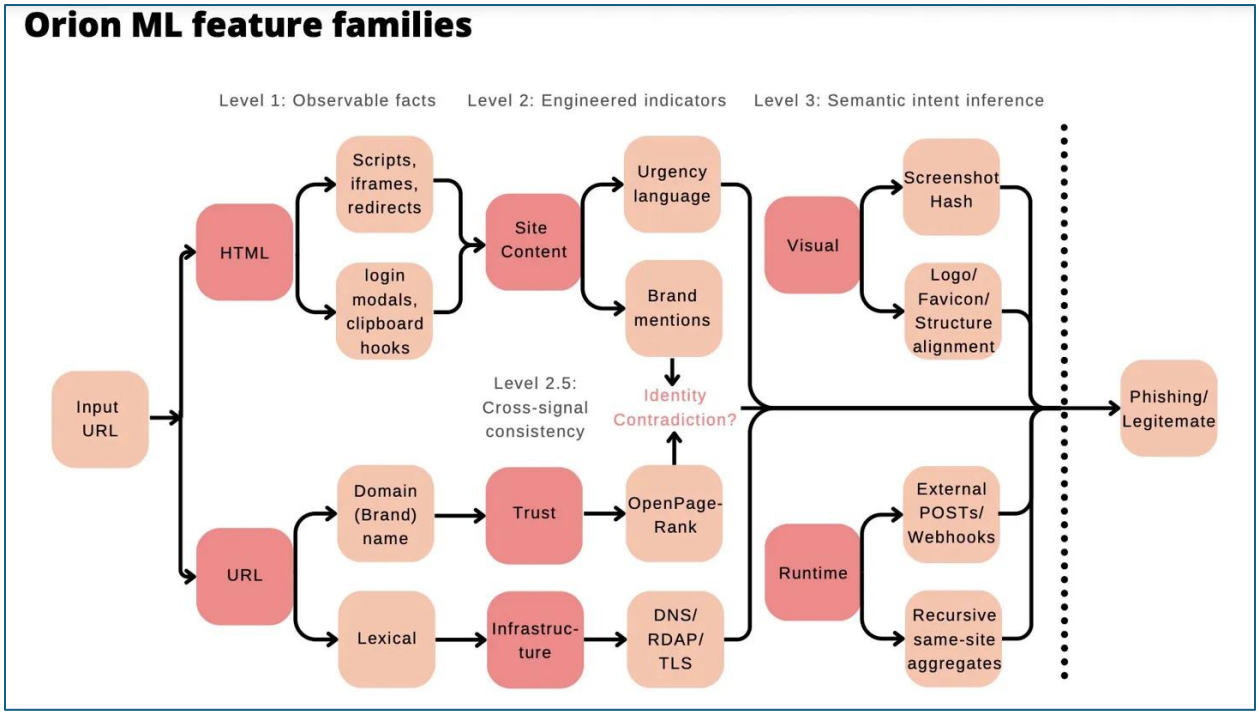
Attackers spend real effort crafting pages that look polished, feel familiar, and arrive at moments when users are distracted or moving quickly. A cloned login page for a service someone uses every day does not trigger instinctive suspicion the way a poorly written email once did. The visual cues that used to signal danger have been steadily erased.

Most organizations respond to this with training programs that were built for a different era. Quarterly simulations, click-through compliance modules, and assigned awareness courses may check a box on an audit, but they rarely change how a user thinks in the moment they are actually targeted. Employees learn what phishing is in the abstract. They rarely learn how to read a specific page in front of them and recognize the signals that give it away.

The gap is not one of awareness. Most users know phishing exists. The gap is one of judgment: the ability to slow down, look at what is actually on the screen, and ask the right questions before typing a password. That kind of judgment is not built in a classroom or a simulation. It is built through repeated, contextual exposure to real evidence. That is the problem Orion was built to solve.

Activities Done to Address the Business Challenge

The problem Orion set out to solve was not just a detection problem. It was a communication problem. Most phishing defense tools stop at the warning: a red flag, a blocked page, a generic alert that tells users something is wrong without explaining why. The goal was to build something that treated the warning as a starting point rather than an endpoint. That meant rethinking the system from the ground up, not just how Orion detects phishing, but how it translates that detection into something a user can actually learn from.

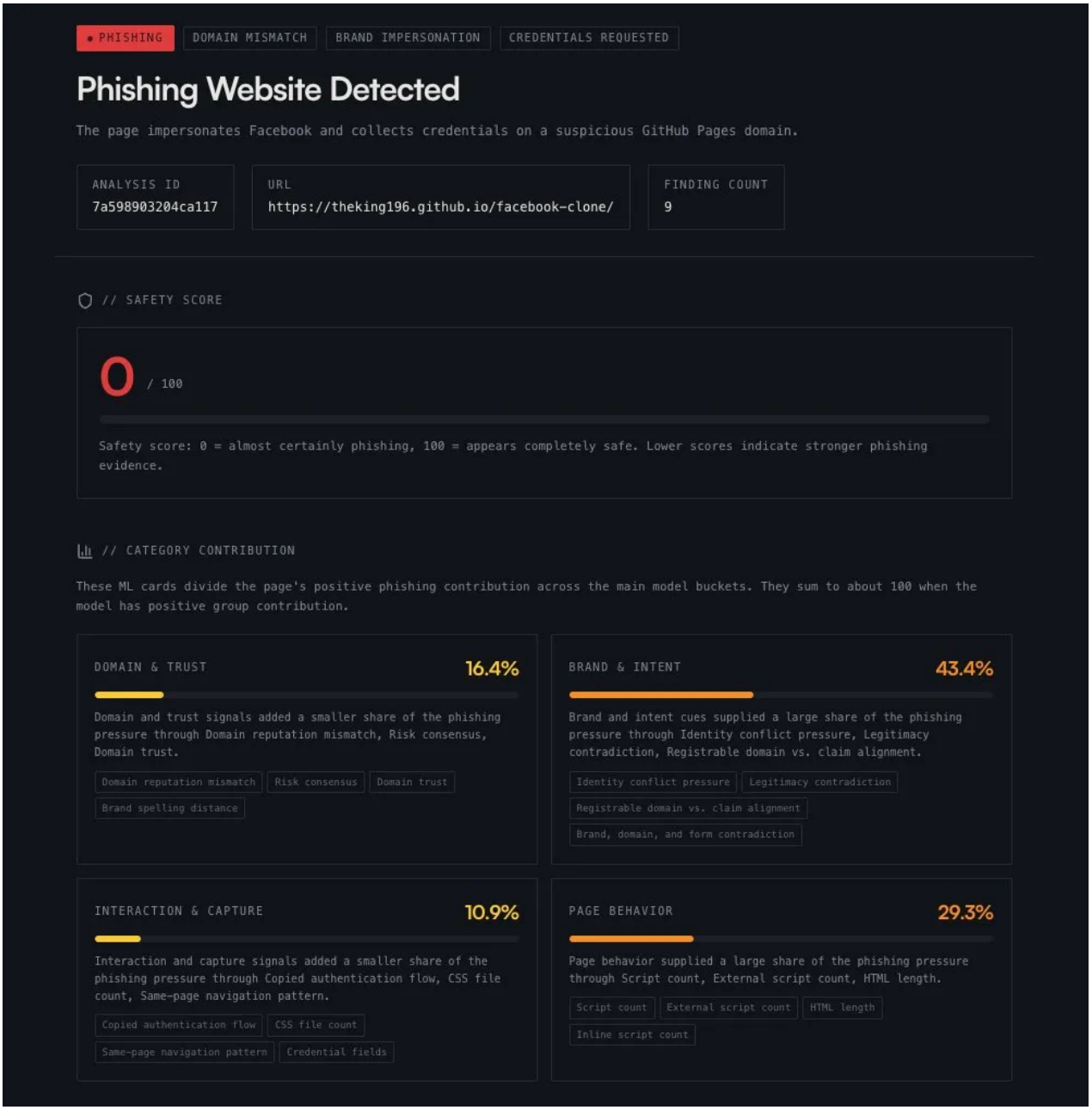


Caption: Chart showing the breakdown of ML feature families.

Orion's detection pipeline is built around a multi-level feature extraction architecture that processes every visited page across two primary input paths: the raw HTML and the URL. From HTML, the system extracts observable facts like the presence of scripts, iframes, redirects, login modals, and clipboard hooks. From the URL, it extracts lexical patterns and the domain's brand name. These observable facts are then engineered into higher-order indicators. HTML feeds into a Site Content layer that surfaces signals like urgency language and brand mentions. The URL feeds into Trust and Infrastructure layers that surface page rank scores, DNS records, RDAP data, and TLS configuration. At Level 3, visual signals are extracted, including screenshot hashes, logo, favicon alignment, and structural layout patterns. Runtime signals are also captured, including external POSTs, webhooks, and recursive same-site aggregates.

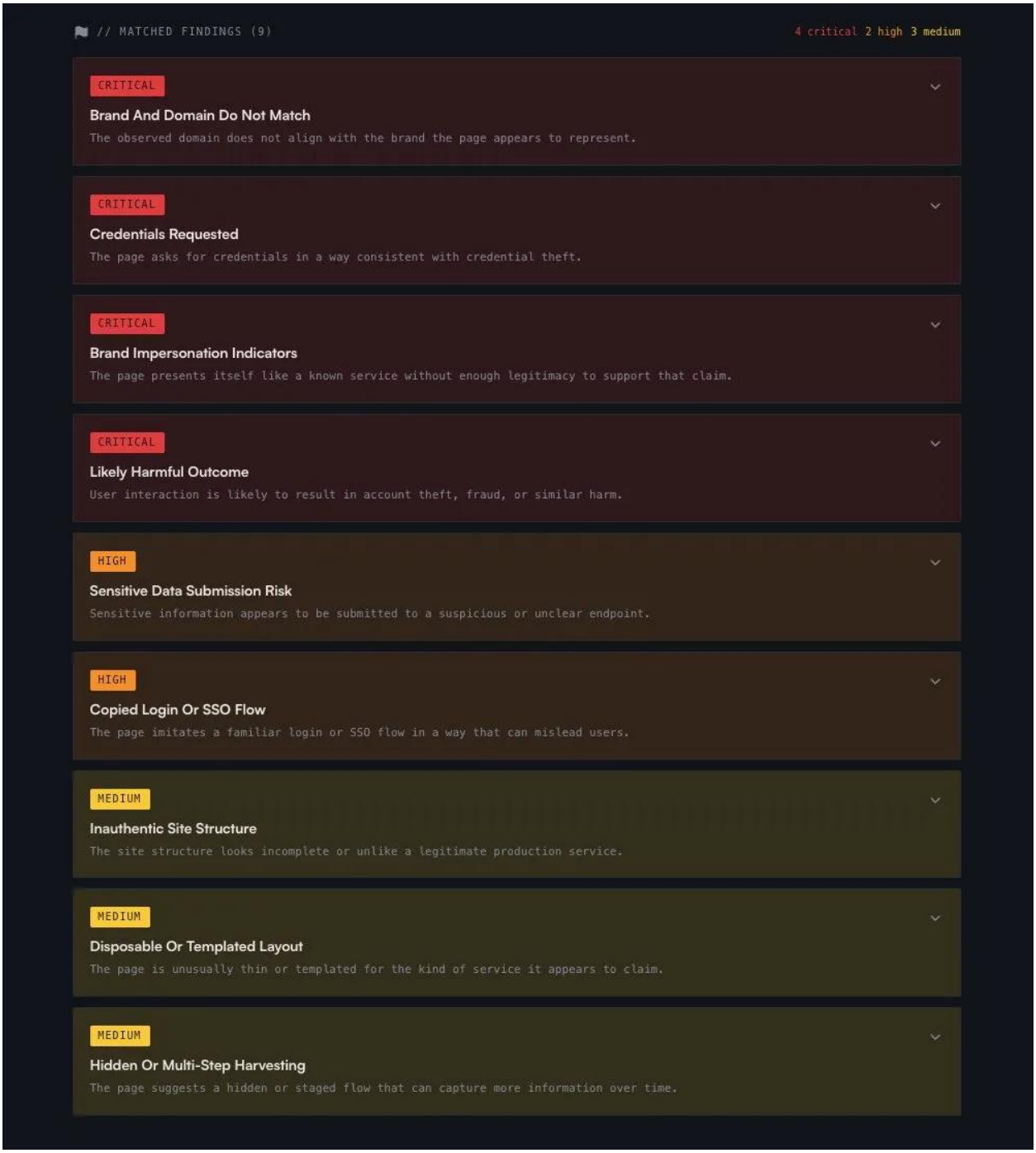
What makes Orion's architecture distinct is what sits between Level 2 and Level 3: a cross-signal consistency check that our team calls Identity Contradiction. Rather than treating each signal in isolation, Orion asks whether the signals across all families tell a coherent story about who the page claims to be. Imagine a page that is displaying a company's logo, using that brand's color palette, and asking for credentials, but is hosted on a domain with no relationship to that brand. That contradiction, when detected, becomes one of the strongest indicators of phishing intent the system can produce.

Once the model produces a raw phishing probability, Orion passes it into an LLM reasoning layer that normalizes the raw signal evidence into a structured set of human-readable findings organized by severity and applies score adjustments based on those findings.



Caption: First section of the guide page that users see when they encounter a phishing site.

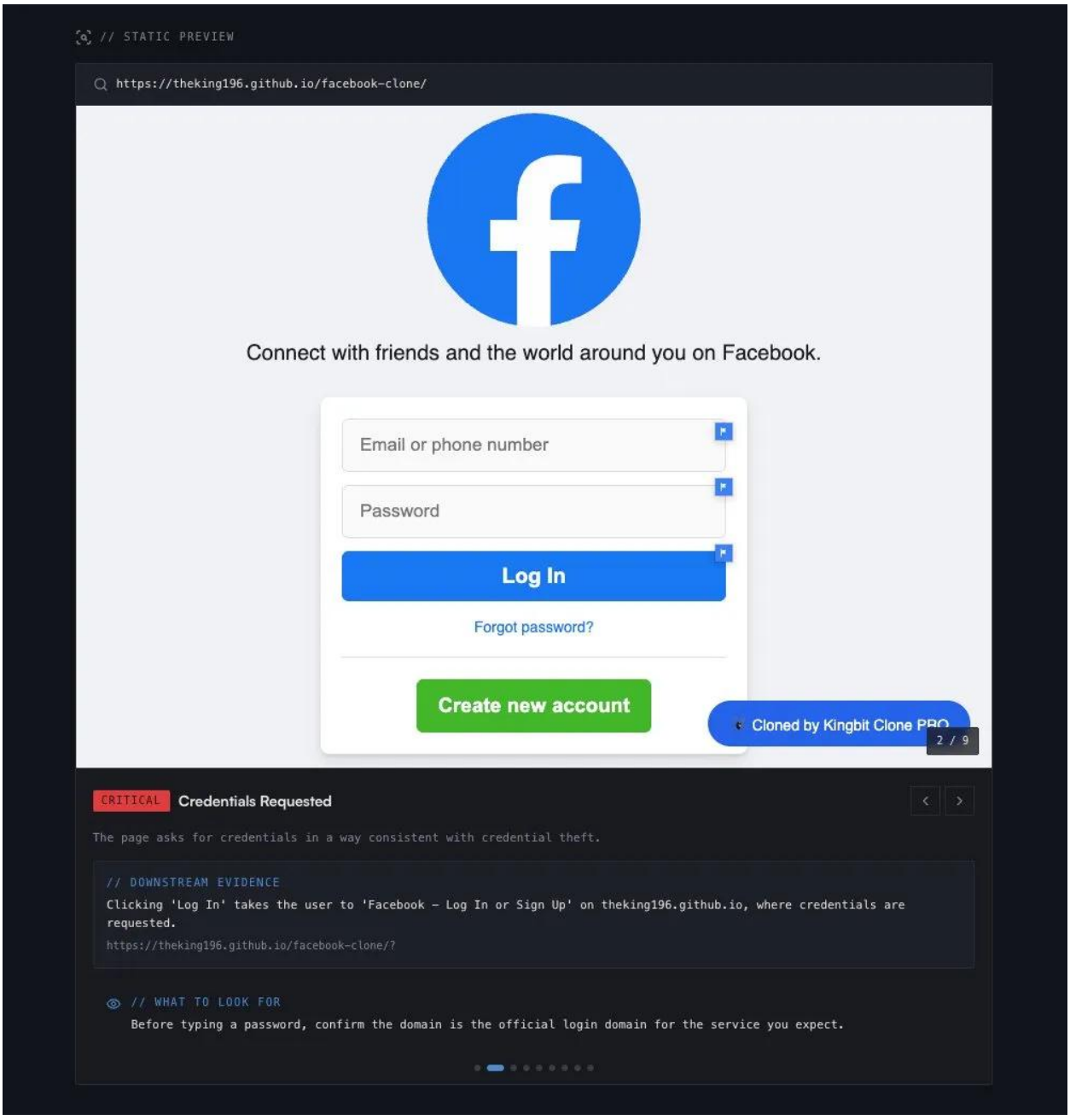
The result is what Orion calls the Safety Score: the final, user-facing number after the ML model's raw phishing probability has been adjusted by the matched findings from the LLM layer. In the example shown, the raw model probability came in at 93.6%. After applying finding-based adjustments, the final Safety Score landed at 0 out of 100, indicating near-certain phishing. Below the score, Orion breaks down the Category Contribution: a percentage view of which ML feature buckets drove the raw phishing signal. In this case, Brand and Intent carried the largest share at 43.4%, followed by Page Behavior at 29.3%, Domain and Trust at 16.4%, and Interaction and Capture at 10.9%. This breakdown gives technically oriented users and administrators visibility into what kinds of signals are most active on any given page.



Caption: Second section of the guide page that users see when they encounter a phishing site.

The matched findings produced by the LLM layer are surfaced as a structured list, each tagged with a severity level and a plain-language description of what the evidence means. In this example, the system returned nine findings: four Critical, two High, and three Medium. The critical findings identified a brand-domain mismatch, credential harvesting behavior, active brand impersonation, and a likely harmful outcome if the user interacted with the page. The

high-severity findings flagged sensitive data submission risk and a copied login or SSO flow. Medium findings noted an inauthentic site structure, a disposable or templated layout, and hidden or multi-step harvesting behavior. Each finding carries a score adjustment that feeds back into the final Safety Score, so the more severe the findings, the lower the score.



Caption: Third section of the guide page that users see when they encounter a phishing site.

Where the system comes into its own is in what happens after those findings are produced. Orion correlates each matched finding to specific elements in a static preview of the flagged page, creating a visual walkthrough that shows users not just what was wrong, but exactly

where on the page the evidence appeared. In the example shown, the page is a Facebook clone hosted on a GitHub Pages domain. Orion's static preview renders the page as the user would have seen it, with findings anchored to the specific DOM elements that triggered them: the credential fields and login button. Navigating through the findings, users see downstream evidence for each one. For the Credentials Requested finding, the system notes that clicking "Log In" routes the user to a credential-harvesting endpoint on a domain with no affiliation to Facebook. Below the evidence, Orion surfaces a plain-language guidance note: before typing a password, confirm the domain is the official login domain for the service you expect.

This is where Orion's core thesis becomes tangible. The guided view is not a report delivered after the fact. It is a real-time teaching moment embedded inside the browsing experience, triggered at the exact second a user would have been most vulnerable. Every finding a user navigates through, every piece of evidence tied to something visible on screen, builds a mental model for what deceptive design looks like. In the future, that guided view can be taken further. Rather than presenting the evidence passively, Orion can require users to identify the suspicious elements themselves, turning exposure into active learning and building the kind of pattern recognition that carries into every future encounter.

Results & The Positive Impact

The research team delivered a working prototype of Orion that demonstrates the full detection and training pipeline from end to end. The system integrates deterministic feature extraction, ML-based inference, and LLM reasoning into a coherent product experience that runs inside the browser without requiring users to change how they work. A user encountering a phishing page receives not just a warning, but a contextualized explanation tied to specific elements on the page in front of them.

The impact is not only in what the system catches. It is in what the system teaches. Each time a user moves through a guided finding, they are building a more accurate mental model of how phishing works in practice. Over repeated encounters, that model compounds. Users who have seen brand-domain mismatches explained clearly are more likely to pause and check the domain bar the next time something feels off, even without Orion present to prompt them.

For organizations, this represents a meaningful shift away from compliance-driven awareness programs toward defense that is embedded in the tools employees already use every day. The project also establishes a foundation for the next phase of development: interactive training exercises that require users to identify suspicious elements rather than simply read about them, an administrative dashboard that surfaces threat trends and training completion across a workforce, and enterprise distribution through managed endpoint platforms that can reach users at scale.

Conclusion

Orion was built on a straightforward conviction: that detection without explanation is a missed opportunity. Every phishing page a system catches is also a chance to leave a user better prepared for the next one. Combining multi-signal analysis, cross-signal identity contradiction,

and LLM-powered explanation into a single closed-loop system allowed the team to pursue both goals at once rather than treating them as separate problems for separate programs.

What the research demonstrated is that this kind of system is buildable. The pipeline works. The guided training view works. The gap between catching a threat and helping a user understand it is one that technology can close, given the right architecture and the right intent behind it. The collaboration across George Mason University, Virginia Tech, Parsons, and the Commonwealth Cyber Initiative made that possible, and the work done here lays the groundwork for continued development toward a platform that can operate at enterprise scale.

PREP Student Reflection

Working on Orion taught us something that no classroom assignment had quite surfaced before: that building a tool and building understanding in the people who use it are two different problems, and that the second one is harder.

We came into this project focused on the technical side. Feature extraction, model accuracy, pipeline architecture. Those were the problems we knew how to frame. But as the system took shape, a more important question kept coming up: what does a user actually walk away with after Orion flags a page? A score and a warning are not enough. We wanted something that left users more capable the next time, with or without Orion in front of them. That question reshaped how we thought about every design decision.

The PREP program gave us the structure and the relationships to pursue that question seriously. The mentorship from Nathan Dykas and Daniel Boyce at Parsons kept the work grounded in how enterprise security tools actually get used: by busy people, under time pressure, with limited tolerance for friction. Brian Ngac, PhD, pushed us to think critically about what we were building and who it was really for. Without that guidance, it would have been easy to optimize for the demo and lose sight of the user.

What we are taking from this is a clearer sense of what it means to build something that genuinely helps people. Technical rigor matters. So does the ability to communicate what a system knows in a way that builds judgment rather than dependence. This project gave us both, and the PREP program made the space for us to find that out together.