# استفاده از مدل‌های زبانی بزرگ برای تشخیص و پاسخ به تهدیدات سایبری: چالش‌ها و راه‌حل‌ها

محمدکاظم سمیعی[۱]، فاطمه حسنی[۲]

[۱] دانشجوی دکتری، گروه علمی مدیریت راهبردی فضای سایبر، دانشگاه و پژوهشگاه عالی دفاع ملی و تحقیقات راهبردی، تهران، اداره کل پدافندغیرعامل
samiei.mk@gmail.com

[۲] دانشجوی کارشناسی ارشد، گروه علمی مهندسی کامپیوتر، هوش مصنوعی و رباتیکز، دانشگاه صنعتی اصفهان، دانشگاه علوم پزشکی قم
mhassani103@gmail.com

**چکیده**

استفاده از مدل‌های زبانی بزرگ (LLMs) مانند GPT-4 و BERT در حوزه امنیت سایبری به‌سرعت در حال گسترش است. این مدل‌ها به دلیل توانایی بالای خود در تحلیل و فهم متون پیچیده، قابلیت‌های جدیدی برای تشخیص و پاسخ به تهدیدات سایبری ارائه می‌دهند. در این مقاله، به بررسی کاربرد مدل‌های زبانی بزرگ برای تشخیص و پاسخ به تهدیدات سایبری پرداخته و چالش‌ها و راه‌حل‌های موجود در این زمینه را تحلیل می‌کنیم. ابتدا مفاهیم اصلی و متدولوژی‌های مورداستفاده را معرفی کرده، سپس به بررسی پیشینه تحقیقاتی و مقایسه با کارهای انجام شده قبلی می‌پردازیم. در نهایت، چالش‌های موجود در استفاده از این مدل‌ها را بررسی کرده و راه‌حل‌های ممکن برای بهبود عملکرد آنها را ارائه می‌دهیم.

**کلمات کلیدی:** مدل‌های زبانی بزرگ، امنیت سایبری، هوش مصنوعی، پردازش زبان طبیعی.

## ۱ مقدمه

در عصر دیجیتال، تهدیدات سایبری به سرعت در حال افزایش و پیچیده‌تر شدن هستند. این تهدیدات شامل حملات فیشینگ، بدافزارها، حملات DDoS و نقض داده‌ها است که می‌توانند خسارات جدی به سازمان‌ها و کاربران وارد کنند. روش‌های سنتی تشخیص و پاسخ به این تهدیدات اغلب ناکارآمد و زمان‌بر هستند و نمی‌توانند به موقع و به صورت کارآمد از وقوع خسارات جلوگیری کنند. بنابراین، نیاز به راهکارهای پیشرفته‌تر و هوشمندانه‌تر بیش از پیش احساس می‌شود.

چالش‌های اصلی در حوزه امنیت سایبری شامل تشخیص به موقع و دقیق تهدیدات، پاسخگویی سریع و مؤثر به حملات، و مدیریت حجم عظیم داده‌های امنیتی است. روش‌های سنتی قادر به پاسخگویی به این

نیازها نیستند و این امر ضرورت بهره‌گیری از فناوری‌های جدیدتر و پیشرفته‌تر را ایجاب می‌کند. هدف این تحقیق بررسی کاربرد مدل‌های زبانی بزرگ (LLMs) مانند GPT-4 و BERT در تشخیص و پاسخ به تهدیدات سایبری است. این مدل‌ها به دلیل توانایی‌های پیشرفته خود در پردازش زبان طبیعی (NLP)، می‌توانند به عنوان ابزارهای قدرتمند در حوزه امنیت سایبری مورد استفاده قرار گیرند. تحقیق حاضر به دنبال شناسایی قابلیت‌ها و محدودیت‌های این مدل‌ها در مقابله با تهدیدات سایبری و ارائه راه‌حل‌هایی برای بهبود کارایی آنها است. این مقاله ابتدا به معرفی مفاهیم اصلی مرتبط با مدل‌های زبانی بزرگ و نحوه کاربرد آنها در امنیت سایبری می‌پردازد. سپس، متدولوژی‌های مورد استفاده برای پیاده‌سازی این مدل‌ها را بررسی می‌کند. در ادامه، با مرور پیشینه تحقیقاتی موجود، به مقایسه کارهای انجام شده قبلی پرداخته و نقاط قوت و ضعف آنها را تحلیل می‌کند. پس از آن، چالش‌های موجود در استفاده از مدل‌های زبانی بزرگ در تشخیص و پاسخ به تهدیدات سایبری را مورد بررسی قرار داده و راه‌حل‌های ممکن برای بهبود عملکرد آنها را ارائه می‌دهد. در نهایت، نتیجه‌گیری و پیشنهاداتی برای تحقیقات آینده ارائه خواهد شد.

## ۲ مفاهیم اصلی

در این بخش، به معرفی مفاهیم کلیدی مرتبط با مدل‌های زبانی بزرگ و کاربرد آنها در حوزه امنیت سایبری می‌پردازیم. این مفاهیم شامل مدل‌های زبانی بزرگ (LLMs)، پردازش زبان طبیعی (NLP)، و کاربردهای این مدل‌ها در تشخیص و پاسخ به تهدیدات سایبری است.

### ۱.۲ مدل‌های زبانی بزرگ (LLMs)

مدل‌های زبانی بزرگ (Large Language Models) مانند GPT-4 و BERT، توسط شرکت‌هایی نظیر OpenAI و Google توسعه داده شده‌اند. این مدل‌ها از تکنیک‌های پیشرفته یادگیری عمیق، برای درک و تولید زبان طبیعی استفاده می‌کنند. LLMها با استفاده از مقادیر عظیمی از داده‌های متنی، آموزش می‌بینند و به صورت خودکار، الگوها و روابط معنایی پیچیده را شناسایی می‌کنند. ویژگی‌های کلیدی این مدل‌ها شامل توانایی درک متن، تولید پاسخ‌های معنی‌دار و شناسایی الگوهای زبانی پیچیده است. مدل‌های زبانی بزرگ می‌توانند از روش‌های مختلفی برای یادگیری استفاده کنند، از جمله:

- یادگیری نظارت‌شده (Supervised Learning): در این روش، مدل با استفاده از مجموعه داده‌های برچسب‌دار آموزش می‌بینند.
- یادگیری نظارت‌نشده (Unsupervised Learning): در این روش، مدل بدون استفاده از برچسب‌ها، از داده‌ها برای یافتن الگوهای پنهان استفاده می‌کند.
- یادگیری انتقالی (Transfer Learning): در این روش، مدل ابتدا بر روی یک مجموعه داده بزرگ آموزش می‌بیند و سپس برای وظایف خاص بهینه‌سازی می‌شود [۱].

## ۲.۲ پردازش زبان طبیعی (NLP)

پردازش زبان طبیعی (Natural Language Processing) شاخه‌ای از هوش مصنوعی است که به تعامل بین کامپیوترها و زبان انسانی می‌پردازد. NLP شامل تکنیک‌هایی است که به کامپیوترها اجازه می‌دهد زبان طبیعی را درک، تفسیر و تولید کنند. تکنیک‌های اصلی NLP شامل موارد زیر است:

- **تجزیه و تحلیل نحوی (Syntactic Analysis)**: شامل تجزیه جمله‌ها به اجزای نحوی مانند فاعل، مفعول و فعل است.
- **تجزیه و تحلیل معنایی (Semantic Analysis)**: شامل تفسیر معنای دقیق کلمات و جملات در متن است.
- **استخراج اطلاعات (Information Extraction)**: شامل استخراج داده‌های ساختاریافته از متون غیرساختاریافته است.
- **تشخیص موجودیت‌های نام‌دار (Named Entity Recognition)**: شامل شناسایی و طبقه‌بندی موجودیت‌های خاص مانند نام‌ها، مکان‌ها و سازمان‌ها در متن است.
- تولید خودکار متن(Text Generation) : شامل تولید متن‌های طبیعی و روان بر اساس داده‌های ورودی است [۲].

## ۳.۲ کاربردهای مدل‌های زبانی بزرگ در امنیت سایبری

مدل‌های زبانی بزرگ به دلیل قابلیت‌های پیشرفته خود در تحلیل متون، کاربردهای گسترده‌ای در حوزه امنیت سایبری دارند. در ادامه، به بررسی برخی از این کاربردها می‌پردازیم:

۱. **تشخیص حملات فیشینگ:**

- **تحلیل ایمیل‌ها و پیام‌های متنی:** مدل‌های زبانی بزرگ می‌توانند محتوای ایمیل‌ها و پیام‌های متنی را تحلیل کرده و الگوهای مشکوک را شناسایی کنند. با استفاده از تکنیک‌های NLP، مدل‌ها قادر به شناسایی نشانه‌های فیشینگ مانند لینک‌های مشکوک، درخواست‌های اطلاعات حساس، و زبان تهدیدآمیز هستند.
- **استخراج اطلاعات حساس:** مدل‌ها می‌توانند اطلاعات حساس مانند شماره‌های حساب، کلمات عبور، و اطلاعات شخصی را شناسایی کرده و از افشای آنها جلوگیری کنند.

۲. **تحلیل لاگ‌های امنیتی:**

- **تشخیص ناهنجاری‌ها:** مدل‌های زبانی بزرگ می‌توانند لاگ‌های امنیتی را تحلیل کرده و الگوهای غیرمعمول یا نشانه‌های حمله را شناسایی کنند. این مدل‌ها قادر به شناسایی ناهنجاری‌ها در رفتار کاربران و سیستم‌ها هستند که می‌تواند نشانه‌ای از یک حمله سایبری باشد.
- **پیش‌بینی حملات:** با تحلیل داده‌های تاریخی، مدل‌های زبانی بزرگ می‌توانند الگوهای حمله را شناسایی کرده و حملات آینده را پیش‌بینی کنند.

۳. **پاسخگویی به تهدیدات:**

- **چت‌بات‌های امنیتی:** مدل‌های زبانی بزرگ می‌توانند به عنوان چت‌بات‌های امنیتی عمل کرده و به سوالات کاربران در مورد تهدیدات سایبری پاسخ دهند. این چت‌بات‌ها می‌توانند راهنمایی‌های لازم برای مقابله با تهدیدات را به کاربران ارائه دهند.
- **پاسخ خودکار به تهدیدات:** مدل‌های زبانی بزرگ می‌توانند به صورت خودکار به تهدیدات سایبری پاسخ دهند. به عنوان مثال، این مدل‌ها می‌توانند دسترسی‌های مشکوک را مسدود کرده یا هشدارهایی را به تیم امنیتی ارسال کنند.

۴. **تحلیل داده‌های شبکه:**

- **شناسایی الگوهای مشکوک:** مدل‌های زبانی بزرگ می‌توانند داده‌های شبکه را تحلیل کرده و الگوهای مشکوک را شناسایی کنند. این تحلیل‌ها می‌توانند به شناسایی و جلوگیری از حملات DDoS و سایر تهدیدات شبکه کمک کنند.
- **تجزیه و تحلیل ترافیک شبکه:** مدل‌ها می‌توانند ترافیک شبکه را به صورت بلادرنگ تحلیل کرده و هرگونه فعالیت غیرمعمول یا مشکوک را شناسایی کنند [۳].

# ۳ متدولوژی

در این بخش، به بررسی متدولوژی‌ها و روش‌های مورد استفاده برای پیاده‌سازی مدل‌های زبانی بزرگ در تشخیص و پاسخ به تهدیدات سایبری می‌پردازیم. این متدولوژی‌ها شامل فرآیندهای جمع‌آوری داده، پیش‌پردازش داده، آموزش مدل، و ارزیابی مدل است.

## ۱.۳ جمع‌آوری داده‌ها

جمع‌آوری داده‌های باکیفیت و متنوع یکی از مراحل کلیدی در پیاده‌سازی مدل‌های زبانی بزرگ است. داده‌های جمع‌آوری شده باید نمایانگر انواع مختلف تهدیدات سایبری و شرایط واقعی باشند. مراحل جمع‌آوری داده شامل موارد زیر است:

۱. **منابع داده:**

- **ایمیل‌ها و پیام‌های متنی:** ایمیل‌ها و پیام‌هایی که ممکن است شامل حملات فیشینگ یا دیگر تهدیدات اجتماعی باشند.
- **لاگ‌های امنیتی:** لاگ‌های سرورها، فایروال‌ها، و سیستم‌های نظارتی که می‌توانند شامل رکوردهای فعالیت‌های مشکوک یا حملات سایبری باشند.
- **داده‌های ترافیک شبکه:** بسته‌های داده و الگوهای ترافیک شبکه که می‌توانند شامل حملات DDoS و دیگر تهدیدات شبکه‌ای باشند.
- **گزارش‌های حملات سایبری:** گزارش‌های مربوط به انواع مختلف حملات سایبری که می‌تواند شامل تکنیک‌ها و روش‌های مختلف حمله باشد.

۲. **فرآیند جمع‌آوری:**

- اتصال به منابع داده: استفاده از ابزارها و تکنیک‌های مختلف برای اتصال به منابع داده و جمع‌آوری اطلاعات.
- مدیریت داده: سازمان‌دهی و ذخیره‌سازی داده‌های جمع‌آوری‌شده به‌گونه‌ای که به‌راحتی قابل دسترسی و تحلیل باشند.

## ۲.۳ پیش‌پردازش داده

پیش‌پردازش داده‌ها شامل مراحل زیر است که به بهبود کیفیت داده‌ها و افزایش دقت مدل‌ها کمک می‌کند:

۱. **پاکسازی داده‌ها:**

- حذف نویزها و داده‌های ناقص: شناسایی و حذف داده‌های ناقص، تکراری، یا نامربوط.
- تصحیح خطاها: شناسایی و تصحیح اشتباهات و ناسازگاری‌ها در داده‌ها.

۲. **نرمال‌سازی داده‌ها:**

- تبدیل به فرم استاندارد: تبدیل داده‌ها به فرم‌های استاندارد برای تضمین یکنواختی و سازگاری.
- مقیاس‌دهی ویژگی‌ها: مقیاس‌دهی ویژگی‌های عددی به بازه‌های معین برای بهبود کارایی مدل.

۳. **توکن‌سازی:**

- تقسیم متن به واحدهای کوچکتر: تقسیم متن به کلمات، عبارات، یا نشانه‌ها برای تحلیل بهتر.

- استفاده از روش‌های توکن‌سازی: استفاده از روش‌های توکن‌سازی مانند BPE (Byte Pair Encoding) یا WordPiece.

۴. **حذف کلمات بی‌اهمیت:**

- استفاده از تکنیک‌های فیلتراسیون: حذف کلمات پرتکرار و بی‌اهمیت که تأثیر کمی بر معنا دارند.
- استفاده از لیست‌های توقف (Stop Words): حذف کلمات توقفی که در تحلیل معنایی کمکی نمی‌کنند.

۵. **پایه‌سازی و لماتیزاسیون:**

- پایه‌سازی (Stemming): کاهش کلمات به ریشه‌های پایه‌ای آنها برای کاهش تنوع زبانی.
- لماتیزاسیون (Lemmatization): تبدیل کلمات به شکل اصلی و معنادار آنها برای دقت بیشتر.

## ۳.۳ آموزش مدل

آموزش مدل‌های زبانی بزرگ یکی از مراحل مهم و پیچیده است که شامل موارد زیر است:

۱. **انتخاب مدل:**

- انتخاب مدل مناسب: انتخاب مدل‌هایی مانند GPT-4 یا BERT بر اساس نیازهای خاص پروژه و کاربرد مورد نظر.
- پیکربندی مدل: تنظیم پیکربندی‌های مدل شامل تعداد لایه‌ها، ابعاد پنهان، و تعداد واحدهای عصبی.

۲. **تنظیمات مدل:**

- تنظیم پارامترهای آموزش: انتخاب پارامترهایی مانند نرخ یادگیری، اندازه دسته، و تعداد دوره‌های آموزش.
- تنظیمات پیشرفته: استفاده از تکنیک‌های پیشرفته مانند Dropout و Regularization برای جلوگیری از Overfitting.

۳. **آموزش مدل:**

- تفکیک داده‌ها: تقسیم داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمون برای ارزیابی بهتر.

- فرآیند آموزش: اجرای فرآیند آموزش با استفاده از الگوریتم‌های یادگیری عمیق مانند Adam یا SGD.
- بهینه‌سازی مدل: استفاده از تکنیک‌های بهینه‌سازی برای تنظیم پارامترهای مدل و بهبود دقت.

## ۴.۳ ارزیابی مدل

ارزیابی مدل برای سنجش دقت و کارایی آن ضروری است و شامل بخش‌های زیر است:

۱. **مجموعه داده‌های آزمون:**

- انتخاب داده‌های آزمون: استفاده از مجموعه داده‌های جداگانه برای ارزیابی عملکرد مدل.
- آزمون‌های بلادرنگ: ارزیابی مدل در شرایط واقعی برای بررسی دقت و کارایی آن.

۲. **معیارهای ارزیابی:**

- دقت (Accuracy): نسبت پیش‌بینی‌های صحیح به کل پیش‌بینی‌ها.
- بازشناسی (Recall): توانایی مدل در شناسایی موارد مثبت واقعی.
- صحت (Precision): درصد موارد مثبت پیش‌بینی شده که واقعاً مثبت هستند.
- F1-score: میانگین هماهنگ دقت و بازشناسی برای ارزیابی کلی مدل.

۳. **آزمون‌های بلادرنگ:**

- پایش عملکرد: نظارت بر عملکرد مدل در شرایط واقعی و جمع‌آوری بازخورد برای بهبود مستمر.
- تصحیح اشتباهات: شناسایی و اصلاح اشتباهات و ناکامی‌ها در مدل برای بهبود عملکرد.

## ۵.۳ بهبود مستمر

برای تضمین عملکرد بهینه مدل‌های زبانی بزرگ، نیاز به بهبود مستمر وجود دارد:

۱. **بازآموزی مدل:**

- استفاده از داده‌های جدید: به‌روزرسانی مدل با استفاده از داده‌های جدید برای بهبود دقت و عملکرد.
- آموزش مجدد: اجرای آموزش مجدد مدل با داده‌های جدید برای افزایش دقت.

۲. **توسعه ویژگی‌های جدید:**

- افزودن ویژگی‌های جدید: توسعه و افزودن ویژگی‌های جدید برای بهبود عملکرد مدل.
- استفاده از تکنیک‌های نوین: پیاده‌سازی تکنیک‌های جدید و پیشرفته برای بهبود عملکرد مدل.

۳. **ارزیابی مستمر:**

- پایش و نظارت: نظارت مداوم بر عملکرد مدل و انجام ارزیابی‌های دوره‌ای.
- بهبود بر اساس بازخورد: استفاده از بازخوردها برای بهبود مستمر و رفع نقاط ضعف مدل [۴].

# ۴ پیشینه تحقیق

در این بخش، به بررسی پیشینه تحقیقاتی و توسعه‌های مرتبط با استفاده از مدل‌های زبانی بزرگ در امنیت سایبری می‌پردازیم. این بررسی شامل مروری بر تحقیقات پیشین، پیشرفت‌ها و چالش‌های موجود در این حوزه است.

## ۱.۴ تاریخچه و پیشرفت‌های اولیه

مدل‌های زبانی بزرگ (LLMs) و تکنیک‌های پردازش زبان طبیعی (NLP) از اواخر دهه ۲۰۱۰ به طور قابل توجهی پیشرفت کرده‌اند. در این دوره، مدل‌های زبانی به مرور زمان از تکنیک‌های ابتدایی مانند مدل‌های شبکه عصبی (Neural Networks) به مدل‌های پیچیده‌تری مانند Transformers تبدیل شدند که توانایی پردازش و تولید متن را به طور چشمگیری بهبود بخشیدند.

- **مدل‌های اولیه NLP:** قبل از ظهور مدل‌های زبانی بزرگ، مدل‌های NLP مانند n-gram models و روش‌های مبتنی بر Bag of Words (BoW) برای تحلیل متن و استخراج ویژگی‌ها استفاده می‌شدند. این مدل‌ها به دلیل محدودیت‌هایشان در درک معنای عمیق متن و تحلیل‌های پیچیده، به مرور زمان جای خود را به مدل‌های پیشرفته‌تری دادند.
- **ظهور مدل‌های Transformers :** در سال ۲۰۱۷، مقاله "Attention is All You Need" که مدل Transformer را معرفی کرد، تحولی بزرگ در پردازش زبان طبیعی به وجود آورد. این مدل به دلیل توانایی‌های خود در پردازش موازی و یادگیری روابط معنایی پیچیده، به سرعت در زمینه‌های مختلف NLP مورد استفاده قرار گرفت [۲].

## ۲.۴ توسعه‌های اخیر در مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ مانند GPT-3 و GPT-4 توسط OpenAI و BERT و T5 توسط Google به عنوان پیشرفت‌های قابل توجه در پردازش زبان طبیعی معرفی شدند. این مدل‌ها توانایی‌های چشمگیری در تولید و تحلیل متن دارند و به طور گسترده‌ای در حوزه‌های مختلفی از جمله امنیت سایبری مورد استفاده قرار گرفته‌اند.

- **GPT-3 و GPT-4 :** مدل‌های Generative Pre-trained Transformer (GPT) از OpenAI، به دلیل قدرت پردازش بالا و توانایی تولید متن‌های بسیار طبیعی، به یکی از ابزارهای اصلی در تحلیل و تولید متن تبدیل شده‌اند. این مدل‌ها با استفاده از تعداد بسیار زیادی پارامتر (مانند ۱۷۵ میلیارد پارامتر در GPT-3) توانایی‌های برجسته‌ای در درک و تولید زبان طبیعی دارند.
- **BERT و T5 :** مدل‌های Bidirectional Encoder Representations from Transformers (BERT) و Text-to-Text Transfer Transformer (T5) از Google نیز به خاطر توانایی‌های خود در پردازش و درک معنای متنی به صورت bidirectional و تولید متون به فرمت‌های مختلف، مورد توجه قرار گرفته‌اند [۱].

## ۳.۴ چالش‌های موجود

در این بخش، به بررسی چالش‌های اصلی مرتبط با استفاده از مدل‌های زبانی بزرگ در تشخیص و پاسخ به تهدیدات سایبری می‌پردازیم. این چالش‌ها شامل مسائل فنی، عملیاتی و اخلاقی هستند که ممکن است در پیاده‌سازی و استفاده از این مدل‌ها بروز کنند [۱۰].

۱. **مشکلات مقیاس‌پذیری و منابع محاسباتی:**

مدل‌های زبانی بزرگ به دلیل تعداد زیاد پارامترها و پیچیدگی ساختاری، نیاز به منابع محاسباتی قابل توجهی دارند. این چالش‌ها شامل موارد زیر است:

- **هزینه‌های بالا:** آموزش و اجرای مدل‌های زبانی بزرگ به منابع سخت‌افزاری گران‌قیمتی نیاز دارد، که می‌تواند برای سازمان‌های کوچک و متوسط چالش‌برانگیز باشد.
- **زمان آموزش:** فرآیند آموزش مدل‌های زبانی بزرگ ممکن است ساعت‌ها یا حتی روزها طول بکشد، که می‌تواند به تأخیر در توسعه و پیاده‌سازی مدل‌ها منجر شود.
- **مدیریت منابع:** نیاز به مدیریت منابع محاسباتی بهینه برای جلوگیری از مصرف بی‌رویه انرژی و افزایش هزینه‌ها.

۲. **دقت و کارایی مدل‌ها:**

مدل‌های زبانی بزرگ ممکن است در برخی شرایط نتوانند به دقت و کارایی مطلوب دست یابند. چالش‌ها شامل موارد زیر است:

- **تنظیمات خاص:** برای بهبود دقت مدل‌ها در تشخیص تهدیدات خاص، نیاز به تنظیمات و بهینه‌سازی‌های خاصی وجود دارد که ممکن است پیچیده و زمان‌بر باشد.
- **مشکلات تعمیم‌پذیری:** مدل‌ها ممکن است در مواجهه با داده‌های جدید یا غیرمنتظره دقت کمتری نشان دهند و نیاز به آموزش مجدد داشته باشند.
- **تولید false positives و false negatives**: مدل‌ها ممکن است با تولید موارد مثبت کاذب (false positives) یا منفی کاذب (false negatives) مواجه شوند که می‌تواند به تحلیل نادرست تهدیدات منجر شود.

۳. **مسائل امنیتی و حریم خصوصی:**

استفاده از مدل‌های زبانی بزرگ در تحلیل داده‌های حساس ممکن است نگرانی‌هایی در مورد امنیت و حریم خصوصی ایجاد کند:

- **حریم خصوصی داده‌ها:** مدل‌های زبانی بزرگ ممکن است نیاز به دسترسی به داده‌های حساس و شخصی داشته باشند که می‌تواند نگرانی‌هایی در مورد حریم خصوصی ایجاد کند.
- **مقابله با حملات:** مدل‌ها ممکن است هدف حملات متقلبانه قرار گیرند که می‌تواند به فاش شدن اطلاعات یا تحریف نتایج منجر شود.
- **حفاظت از داده‌ها:** تضمین اینکه داده‌های مورد استفاده برای آموزش مدل‌ها به درستی مدیریت و محافظت شوند.

۴. **پیچیدگی و هزینه‌های اجرایی:**

پیاده‌سازی و نگهداری مدل‌های زبانی بزرگ شامل چالش‌های عملیاتی زیر است:

- **پیچیدگی پیاده‌سازی:** پیاده‌سازی مدل‌های زبانی بزرگ نیازمند تخصص‌های فنی و دانش عمیق در زمینه یادگیری عمیق و پردازش زبان طبیعی است.
- **هزینه‌های نگهداری:** نگهداری و به‌روزرسانی مدل‌ها به دلیل نیاز به داده‌های جدید و آموزش مجدد می‌تواند هزینه‌بر باشد.
- **آموزش و پشتیبانی:** نیاز به آموزش کارکنان و ارائه پشتیبانی فنی برای بهره‌برداری مؤثر از مدل‌ها.

۵. **مشکلات اخلاقی و اجتماعی:**

استفاده از مدل‌های زبانی بزرگ در زمینه‌های مختلف می‌تواند مسائل اخلاقی و اجتماعی را به همراه داشته باشد:

- **مسئولیت‌پذیری:** تعیین مسئولیت در صورت بروز خطا یا آسیب ناشی از استفاده نادرست از مدل‌ها.
- **تبعیض و نابرابری:** امکان وجود تبعیض در نتایج مدل‌ها بر اساس داده‌های ورودی و تأثیر آن بر گروه‌های مختلف اجتماعی.
- **تجارت و سودجویی:** استفاده تجاری از مدل‌ها و احتمال سوءاستفاده از آنها برای اهداف نادرست.

۶. **چالش‌های مربوط به داده‌ها:**

مدل‌های زبانی بزرگ به داده‌های گسترده و باکیفیت نیاز دارند که ممکن است شامل چالش‌های زیر باشد:

- **داده‌های ناکافی یا نامناسب:** نبود داده‌های کافی یا نامناسب می‌تواند به دقت پایین مدل‌ها منجر شود.
- **تنوع داده‌ها:** نیاز به داده‌های متنوع برای اطمینان از توانایی مدل‌ها در تعمیم به شرایط مختلف.
- **پیش‌پردازش داده:** فرآیند پیش‌پردازش داده‌ها ممکن است پیچیده و زمان‌بر باشد و نیاز به دقت بالا داشته باشد [۱۱].

## ۴.۴ مقایسه کارهای انجام شده قبلی

در این بخش، به مقایسه تحقیقات و کارهای انجام شده قبلی در زمینه استفاده از مدل‌های زبانی بزرگ برای تشخیص و پاسخ به تهدیدات سایبری می‌پردازیم. این مقایسه شامل بررسی روش‌ها، نتایج و محدودیت‌های تحقیقات مختلف است.

### کارهای انجام شده در تشخیص حملات فیشینگ

۱. **PhishNet**:

در این پژوهش "PhishNet: Detecting Phishing Websites Using Deep Learning" به بررسی استفاده از مدل BERT برای شناسایی حملات فیشینگ پرداخته شده است. این تحقیق نشان داد که BERT به دلیل توانایی‌های خود در درک عمیق متن، قادر به شناسایی الگوهای فیشینگ با دقت بالاست. BERT توانست به طور مؤثر و دقیق لینک‌های فیشینگ را از لینک‌های مشروع تفکیک کند. این تحقیق نشان داد که BERT می‌تواند با استفاده از داده‌های متنی، تهدیدات فیشینگ را با دقت بالا شناسایی کند. یکی از محدودیت‌های این تحقیق نیاز به داده‌های آموزشی گسترده و پیچیدگی محاسباتی بالا بود [۱۲].

۲. **روش مبتنی بر مدل‌های Generative Pre-trained Transformers (GPT):**
در این تحقیق "Detecting Phishing Attacks in Emails Using GPT-3" به بررسی استفاده از GPT-3 برای شناسایی حملات فیشینگ در ایمیل‌ها پرداخته شده است. این تحقیق نشان داد که GPT-3 قادر به تولید و تحلیل متن‌های طبیعی است و می‌تواند به شناسایی الگوهای مشکوک در ایمیل‌ها کمک کند. GPT-3 توانست به طور مؤثری به شناسایی ایمیل‌های فیشینگ کمک کند و نرخ شناسایی بالایی را به ثبت رساند. محدودیت‌های GPT-3 شامل نیاز به منابع محاسباتی زیاد و پیچیدگی‌های ناشی از تولید متن‌های غیرمعمول بود [۱۳].

**کارهای انجام شده در تحلیل لاگ‌های امنیتی**

۱. **روش استفاده از مدل‌های Transformer برای تحلیل لاگ‌ها:**
در این پژوهش "LogNet: Deep Learning for Anomaly Detection in Security Logs" به بررسی استفاده از مدل‌های Transformer برای تحلیل لاگ‌های امنیتی پرداخته شده است. این تحقیق نشان داد که مدل‌های Transformer می‌توانند ناهنجاری‌های موجود در لاگ‌ها را با دقت بالا شناسایی کنند. مدل‌های Transformer قادر به شناسایی الگوهای غیرمعمول در لاگ‌ها بودند و توانایی تحلیل حجم‌های بزرگ داده را داشتند. یکی از محدودیت‌های این تحقیق، پیچیدگی بالا و نیاز به پردازش داده‌های زیاد برای آموزش مدل بود [۱۴].

۲. **بررسی کاربردهای NLP در تحلیل لاگ‌ها:**
در این تحقیق "Using NLP for Log Analysis and Anomaly Detection" به بررسی کاربردهای پردازش زبان طبیعی در تحلیل لاگ‌های امنیتی پرداخته شده است. این تحقیق نشان داد که تکنیک‌های NLP می‌توانند به شناسایی و تحلیل ناهنجاری‌ها کمک کنند. استفاده از تکنیک‌های NLP به شناسایی بهتر و سریع‌تر ناهنجاری‌ها و تهدیدات کمک کرد. چالش‌های اصلی، شامل نیاز به تنظیمات خاص برای داده‌های مختلف و دقت پایین در شناسایی برخی ناهنجاری‌ها بود [۱۵].

**کارهای انجام شده در پاسخگویی به تهدیدات**

۱. **چت‌بات‌های امنیتی مبتنی بر مدل‌های زبانی بزرگ:**
در این تحقیق "Security Chatbots: Analyzing User Queries with Deep Learning Models" به بررسی استفاده از مدل‌های زبانی بزرگ در ایجاد چت‌بات‌های امنیتی پرداخته شده است. این تحقیق به تحلیل قابلیت‌های مدل‌های زبانی بزرگ در پاسخ به سوالات امنیتی کاربران پرداخته است. مدل‌های زبانی بزرگ توانستند به طور مؤثری به سوالات امنیتی پاسخ دهند و اطلاعات مفیدی ارائه دهند. محدودیت‌های این تحقیق شامل توانایی محدود مدل‌ها در درک تمامی پیچیدگی‌های امنیتی و نیاز به به‌روزرسانی مداوم بود [۱۶].

۲. **مدل‌های زبان برای پاسخ خودکار به تهدیدات:**
در این پژوهش "Security Chatbots: Analyzing User Queries with Deep Learning

Models" به بررسی استفاده از مدل‌های زبانی بزرگ برای پاسخ خودکار به تهدیدات پرداخته شده است. این تحقیق نشان داد که مدل‌ها قادر به ارائه پاسخ‌های خودکار و مناسب به تهدیدات بودند. استفاده از مدل‌های زبانی بزرگ برای پاسخ به تهدیدات توانست بهبود قابل توجهی در زمان واکنش به تهدیدات ایجاد کند. چالش‌های اصلی شامل دقت پایین در برخی موارد و نیاز به تنظیمات خاص برای انواع مختلف تهدیدات بود [۱۷].

**نتایج مقایسه**

از بررسی مدل‌های استفاده شده در کارهای انجام شده قبلی (جدول ۱) می‌توان دریافت:

- **BERT :** دقت بالایی در تشخیص تهدیدات سایبری مانند فیشینگ دارد، اما به منابع محاسباتی زیادی نیازمند است و برای داده‌های حساس ممکن است نیاز به تنظیم دقیق داشته باشد.
- **GPT-3:** توانایی تحلیل و پاسخگویی به تهدیدات را دارد، اما به دلیل پیچیدگی و حجم بالای محاسبات، نیازمند زمان و منابع زیادی است.
- **ترکیب BERT و CNN :** می‌تواند به بهبود دقت و تشخیص الگوهای پیچیده کمک کند، اما چالش‌های اجرایی مانند مصرف منابع زیاد و هماهنگی بین مدل‌ها وجود دارد.
- **RoBERTa:** با بهینه‌سازی‌های بیشتر نسبت به BERT، دقت بهتری دارد اما برای داده‌های ناهنجار پیچیده نیاز به آموزش‌های سنگین دارد.
- **T5:** توانایی تشخیص حملات شبکه‌ای مانند DDoS را دارد، اما به تنظیمات دقیق و منابع گسترده نیاز دارد.
- **BERT: Multilingual** برای تحلیل تهدیدات چندزبانه مناسب است اما برای زبان‌های کمتر استفاده‌شده ممکن است دقت کمتری داشته باشد.
- **DistilBERT:** با کاهش حجم و منابع مصرفی، سرعت و کارایی بالاتری دارد، اما ممکن است دقت کمتری نسبت به مدل‌های بزرگتر داشته باشد.

**خلاصه بررسی‌ها**

در مجموع، تحقیقات نشان داده‌اند که مدل‌های زبانی بزرگ دارای توانایی‌های قابل توجهی در زمینه‌های مختلف امنیت سایبری هستند. استفاده از این مدل‌ها می‌تواند به بهبود دقت و کارایی در تشخیص و پاسخ به تهدیدات سایبری کمک کند. با این حال، چالش‌هایی مانند نیاز به منابع محاسباتی بالا، پیچیدگی‌های اجرایی و محدودیت‌های دقت همچنان وجود دارد که نیازمند تحقیقات و بهبودهای بیشتر هستند.

جدول ۱: مقایسه کارهای انجام شده قبلی

| کار پژوهشی | مدل | قابلیت‌ها | مزایا | محدودیت‌ها |
|---|---|---|---|---|
| تشخیص حملات فیشینگ با BERT | BERT | تشخیص حملات فیشینگ از طریق تحلیل متن ایمیل‌ها | دقت بالا در تشخیص حملات فیشینگ، توانایی پردازش مقادیر زیاد متن | نیاز به منابع محاسباتی بالا، حساسیت به تغییرات کوچک در داده‌های متنی |
| استفاده از GPT-3در تحلیل لاگ‌های امنیتی | GPT-3 | تحلیل و تفسیر خودکار لاگ‌های امنیتی | توانایی تولید متن خودکار، بهبود سرعت تحلیل لاگ‌های امنیتی | پیچیدگی و زمان طولانی آموزش، نیاز به تنظیم دقیق برای حوزه‌های خاص امنیتی |
| ترکیب BERT و CNN برای تشخیص بدافزار | ترکیب BERT و CNN | تشخیص بدافزار از طریق تحلیل رفتارهای سیستم و فایل‌های مشکوک | ترکیب قدرت تحلیل زبان با تشخیص الگوهای پیچیده، بهبود دقت در تشخیص تهدیدات پیچیده | نیاز به منابع محاسباتی بیشتر برای اجرای ترکیب دو مدل بزرگ، چالش در تنظیم دقیق و هماهنگی بین مدل‌ها |
| استفاده از RoBERTa برای تشخیص ناهنجاری‌های امنیتی | RoBERTa | تحلیل ناهنجاری‌های امنیتی در سیستم‌های پیچیده و شناسایی الگوهای ناهنجار | بهبود دقت نسبت به BERT بهدلیل تنظیمات بهینه‌تر در مدل، مناسب برای داده‌های پیچیده و بدون ساختار | نیاز به داده‌های بیشتر برای آموزش، حساسیت به نوع داده‌های ورودی |
| تشخیص حملات DDoS با T5 | T5 | تشخیص و پاسخ به حملات DDoS از طریق تحلیل ترافیک شبکه | قابلیت تولید پاسخ خودکار به حملات، بهبود دقت در تشخیص تهدیدات شبکه‌ای | نیاز به تنظیم دقیق و استفاده از منابع گسترده برای تحلیل ترافیک شبکه |
| تحلیل متون سایبری با Multilingual BERT | Multilingual BERT | تشخیص تهدیدات سایبری در متون چندزبانه | قابلیت کار با داده‌های چندزبانه، افزایش دقت در تحلیل تهدیدات بین‌المللی | چالش در بهبود دقت برای زبان‌های کمتر پشتیبانی‌شده، نیاز به داده‌های بزرگ برای هر زبان |
| تحلیل لاگ‌های امنیتی با DistilBERT | DistilBERT | تحلیل سریع و کارآمد لاگ‌های امنیتی با مدل فشرده | کاهش مصرف منابع محاسباتی نسبت به BERT، سرعت بالاتر در پردازش داده‌های لاگ | دقت کمتر نسبت به مدل‌های کامل مانند BERT |

# ۵ راهکار پیشنهادی

در این بخش، به بررسی راهکارهای موجود برای حل چالش‌های مرتبط با استفاده از مدل‌های زبانی بزرگ در تشخیص و پاسخ به تهدیدات سایبری می‌پردازیم. راهکارهایی که به بهبود عملکرد و کارایی مدل‌ها کمک می‌کنند، در زیر توضیح داده شده‌اند.

## ۱.۵ بهینه‌سازی منابع محاسباتی

- **استفاده از مدل‌های کم‌حجم و فشرده:** استفاده از نسخه‌های فشرده‌شده مدل‌های بزرگ مانند DistilBERT، TinyBERT، و MobileBERT باعث کاهش نیاز به منابع محاسباتی و حافظه می‌شوند. این مدل‌ها توانایی حفظ دقت مدل‌های بزرگ را در مقیاس کوچکتر دارند. توسعه و استفاده از مدل‌های فشرده اختصاصی برای حوزه‌های خاص امنیت سایبری، مانند مدل‌هایی که به طور خاص برای تشخیص حملات فیشینگ یا تحلیل لاگ‌های امنیتی آموزش داده شده‌اند، پیشنهاد می‌شود [۱۸]، [۱۹]، [۲۰].

- **مقابله با افزایش مقیاس:** بهره‌گیری از خدمات ابری و معماری‌های توزیع‌شده برای مقیاس‌پذیری بهتر از قبیل استفاده از زیرساخت‌های محاسبات ابری مانند AWS Sagemaker، Google AI Platform، و Azure Machine Learning برای اجرای مدل‌ها راهکار مطلوبی در برابر افزایش مقیاس داده‌ها است. در این زمینه ایجاد پلتفرم‌های ابری اختصاصی برای سازمان‌های امنیت سایبری که شامل ابزارهای خاص برای بهینه‌سازی آموزش و اجرای مدل‌های زبانی بزرگ باشد، پیشنهاد می‌شود.

- **بهینه‌سازی محاسبات:** جهت بهینه سازی محاسبات می‌توان تکنیک‌های pruning، quantization، و knowledge distillation برای کاهش حجم مدل و بهبود سرعت اجرا بکارگیری کرد. همچنین توسعه ابزارهای خودکار برای شناسایی و اعمال بهینه‌سازی‌های مناسب برای مدل‌های زبانی بزرگ در زمینه امنیت سایبری توصیه شده است.

## ۲.۵ بهبود دقت و کارایی مدل‌ها

- **تنظیم دقیق مدل‌ها:** برای تنظیم دقیق مدلها می‌توان از تکنیک‌های تنظیم دقیق (fine-tuning) با داده‌های تخصصی و تنظیمات خاص برای بهبود دقت مدل‌ها در زمینه‌های خاص مانند تشخیص تهدیدات سایبری استفاده کرد.

  ایجاد مجموعه‌های داده‌های خاص برای آموزش مدل‌ها در زمینه‌های مختلف امنیت سایبری و بهبود دقت مدل‌ها با استفاده از داده‌های جدید و متنوع نیز می‌تواند موثر باشد.

- **استفاده از تکنیک‌های Ensemble:** ترکیب چندین مدل برای بهبود دقت و کاهش نرخ اشتباهات

از قبیل تکنیک‌های ensemble مانند bagging و boosting می‌توانند به بهبود عملکرد کلی سیستم کمک کنند.

- **استفاده از داده‌های افزوده:** به کارگیری تکنیک‌های داده‌افزایی (data augmentation) مانند ترجمه ماشینی، تغییرات در ساختار متن، و ایجاد داده‌های شبیه‌سازی شده برای افزایش تنوع داده‌ها و توسعه روش‌های داده‌افزایی خاص برای داده‌های امنیت سایبری، شامل ایجاد داده‌های شبیه‌سازی شده برای حملات سایبری و ناهنجاری‌های امنیتی، پیشنهاد می‌شود.

## ۳.۵ مدیریت مسائل امنیتی و حریم خصوصی

- **رمزگذاری و محافظت از داده‌ها:** استفاده از تکنیک‌های رمزگذاری پیشرفته و حفاظت از داده‌ها در فرآیند جمع‌آوری و تحلیل و پیاده‌سازی چارچوب‌های امنیتی خاص برای مدل‌های زبانی بزرگ در زمینه امنیت سایبری، شامل رمزگذاری و حفاظت از داده‌های حساس، می‌تواند ضریب اطمینان لازم را در محافظت از داده‌ها ایجاد نماید.

- **اجرای قوانین حریم خصوصی:** با پیروی از مقررات حریم خصوصی مانند GDPR و CCPA و اجرای سیاست‌های حفظ حریم خصوصی در تمام مراحل استفاده از مدل‌ها و ایجاد سیاست‌های حریم خصوصی و امنیت داده‌های خاص برای مدل‌های زبانی بزرگ که شامل تکنیک‌های خاص برای محافظت از داده‌های امنیت سایبری باشد، می‌توان مسائل مرتب با حریم خصوصی را مدیریت نمود.

- **مراقبت از امنیت مدل:** پیاده‌سازی تدابیر امنیتی برای محافظت از مدل‌ها در برابر حملات و سوءاستفاده و توسعه ابزارهای امنیتی خاص برای مدل‌های زبانی بزرگ شامل نظارت بر فعالیت‌های مشکوک و شناسایی حملات متقلبانه، نقش مهمی در مراقبت از امنیت مدل خواهد داشت.

## ۴.۵ ساده‌سازی پیچیدگی و کاهش هزینه‌های اجرایی

- **استفاده از مدل‌های پیش‌آموزش‌دیده:** بهره‌گیری از مدل‌های زبانی بزرگ که قبلاً آموزش داده شده‌اند و در دسترس هستند، برای کاهش هزینه‌ها و پیچیدگی‌های اجرایی مناسب هستند. همچنین ایجاد مدل‌های پیش‌آموزش‌دیده ی ویژه، برای حوزه‌های مختلف امنیت سایبری با استفاده از داده‌های خاص و نیازهای کاربردی نیز پیشنهاد می‌شود.

- **بهینه‌سازی فرآیند آموزش:** استفاده از الگوریتم‌های پیشرفته یادگیری و کاهش حجم داده‌های مورد نیاز، برای آموزش و توسعه ابزارهای خودکار برای بهینه‌سازی فرآیند آموزش مدل‌ها، در زمینه‌های خاص امنیت سایبری، برای بهینه سازی فرآیند آموزش پیشنهاد می‌شود.

- **آموزش و پشتیبانی:** ارائه برنامه‌های آموزشی و پشتیبانی فنی برای کارکنان به منظور تسهیل در استفاده و نگهداری از مدل‌های زبانی بزرگ و تدوین دوره‌های آموزشی تخصصی و ارائه خدمات

مشاوره فنی، برای پیاده‌سازی و مدیریت مدل‌های زبانی بزرگ در سازمان‌های امنیت سایبری مورد تأکید است.

### ۵.۵ پرداختن به مسائل اخلاقی و اجتماعی

- **توسعه چارچوب‌های اخلاقی:** ایجاد و پیاده‌سازی چارچوب‌های اخلاقی برای استفاده از مدل‌های زبانی بزرگ، شامل اصول شفافیت، انصاف و مسئولیت‌پذیری و توسعه دستورالعمل‌های اخلاقی اختصاصی برای مدل‌های زبانی بزرگ در زمینه امنیت سایبری و تضمین رعایت اصول اخلاقی در تمامی مراحل توسعه و استفاده، مد نظر است.
- **مدیریت تبعیض و نابرابری:** استفاده از تکنیک‌های متوازن‌سازی داده‌ها و ارزیابی مدل برای جلوگیری از تبعیض و نابرابری در نتایج مدل‌ها و پیاده‌سازی روش‌های ارزیابی و اصلاح مدل برای شناسایی و کاهش تبعیض‌های ممکن در تحلیل‌های امنیتی، پیشنهاد می‌شود.

### ۶.۵ حل مشکل داده‌ها

- **افزایش تنوع داده‌ها:** جمع‌آوری و استفاده از داده‌های متنوع برای بهبود توانایی مدل‌ها در تحلیل و پردازش داده‌های مختلف، همچنین ایجاد پایگاه‌های داده تخصصی شامل داده‌های شبیه‌سازی شده و واقعی برای تهدیدات سایبری خاص.
- **بهبود پیش‌پردازش داده‌ها:** استفاده از تکنیک‌های پیش‌پردازش پیشرفته برای بهبود کیفیت داده‌ها و کاهش خطاهای ورودی و توسعه ابزارهای خودکار برای پیش‌پردازش داده‌های امنیت سایبری شامل تحلیل و اصلاح داده‌های متنی و ساختاری.
- **به‌روزرسانی داده‌ها:** به روزرسانی و ارتقاء داده‌های آموزشی به طور مداوم برای حفظ دقت و کارایی مدل‌ها و ایجاد فرآیندهای خودکار برای جمع‌آوری و ادغام داده‌های جدید به مدل‌ها به منظور حفظ دقت و کارایی در برابر تهدیدات جدید.

## ۶ نتیجه‌گیری

در این مقاله، استفاده از مدل‌های زبانی بزرگ برای تشخیص و پاسخ به تهدیدات سایبری مورد بررسی قرار گرفت. این فناوری‌ها، با قابلیت‌های پیشرفته خود در پردازش زبان طبیعی، امکان تحلیل دقیق و سریع داده‌ها را فراهم می‌کنند، اما همچنان با چالش‌های متعددی مواجه هستند.

مدل‌های زبانی بزرگ مانند BERT، GPT و سایر نسخه‌های پیشرفته، توانسته‌اند در زمینه‌های مختلف امنیت سایبری، از جمله تشخیص حملات فیشینگ، تحلیل لاگ‌های امنیتی و پاسخگویی به تهدیدات، عملکرد قابل توجهی را نشان دهند. این مدل‌ها قادرند با دقت بالا به شناسایی تهدیدات و ناهنجاری‌ها کمک کنند و سرعت پاسخ به تهدیدات را بهبود بخشند.

با وجود مزایای فراوان، استفاده از مدل‌های زبانی بزرگ با چالش‌هایی نظیر نیاز به منابع محاسباتی بالا، مسائل دقت و کارایی، نگرانی‌های امنیتی و حریم خصوصی، پیچیدگی‌های اجرایی و مسائل اخلاقی مواجه است. این چالش‌ها می‌تواند بر روی اجرای موفقیت‌آمیز این فناوری‌ها تأثیر بگذارد و نیاز به توجه و حل مشکلات خاص خود را داشته باشد.

برای مقابله با این چالش‌ها، راهکارهای متعددی ارائه شده است. از جمله استفاده از مدل‌های کم‌حجم و فشرده، بهینه‌سازی منابع محاسباتی، تنظیم دقیق مدل‌ها و استفاده از تکنیک‌های ensemble و داده‌افزایی. همچنین، راهکارهای مدیریت مسائل امنیتی و حریم خصوصی، ساده‌سازی پیچیدگی‌های اجرایی، و پرداختن به مسائل اخلاقی و اجتماعی مورد بررسی قرار گرفتند.

همچنین توسعه مدل‌های فشرده ویژه برای امنیت سایبری، ایجاد پایگاه‌های داده تخصصی، و استفاده از ابزارهای خودکار برای بهینه‌سازی آموزش و پیش‌پردازش داده‌ها، پیشنهاد شده است. این راهکارها می‌توانند به بهبود کارایی و کاهش هزینه‌های مرتبط با استفاده از مدل‌های زبانی بزرگ کمک کنند و به پیاده‌سازی مؤثرتر این فناوری‌ها در زمینه امنیت سایبری منجر شوند.

## مراجع

[1] Yao, Yifan, et al. “A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly”. High-Confidence Computing, 2024.

[2] O’Connor, Joseph, and Ian McDermott. Principles of NLP: What it is, how it works. Singing Dragon, 2013. Zhang, Jie, et al. “When LLMs meet cybersecurity: A systematic literature review”. arXiv preprint arXiv:2405.03644, 2024.

[3] Capodieci, Noelle, et al. “The Impact of Generative AI and LLMs on the Cybersecurity Profession”. 2024 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2024.

[4] Yamin, Muhammad Mudassar, et al. “Applications of llms for generating cyber security exercise scenarios”, 2024.

[5] Ruzickova, Miroslava, et al. “AI and LLM Models to Analyze and Identify Cybersecurity Incidents”.

[6] Laney, Samuel P. LLM-Directed Agent Models in Cyberspace. Diss. Massachusetts Institute of Technology, 2024.

[7] Gennari, Jeff, et al. “Considerations for evaluating large language models for cybersecurity tasks”, 2024.

[8] Majumdar, Subhabrata. “Standards for LLM Security”. Large, 225, 2024.

[9] Motlagh, Farzad Nourmohammadzadeh, et al. “Large language models in cybersecurity: State-of-the-art”. arXiv preprint arXiv:2402.00891, 2024.

[10] Ayoobi, Navid, Sadat Shahriar, and Arjun Mukherjee. “The looming threat of fake and llm-generated linkedin profiles: Challenges and opportunities for detection and prevention”. Proceedings of the 34th ACM Conference on Hypertext and Social Media. 2023.

[11] Jiao, Junfeng, et al. “Navigating LLM ethics: Advancements, challenges, and future directions”. arXiv preprint arXiv:2406.18841, 2024.

[12] Prakash, Pawan, et al. “Phishnet: predictive blacklisting to detect phishing attacks”. 2010 Proceedings IEEE INFOCOM. IEEE, 2010.

[13] Patel, Het, Umair Rehman, and Farkhund Iqbal. “Large Language Models Spot Phishing Emails with Surprising Accuracy: A Comparative Analysis of Performance”. arXiv preprint arXiv:2404.15485, 2024.

[14] Lee, Edward H., et al. “Lognet: Energy-efficient neural networks using logarithmic computation”. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.

[15] Bertero, Christophe, et al. “Experience report: Log mining using natural language processing and application to anomaly detection”. 2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2017.

[16] Hasal, Martin, et al. “Chatbots: Security, privacy, data protection, and social aspects”. Concurrency and Computation: Practice and Experience 33.19, 2021.

[17] Sultana, Madeena, et al. “Towards evaluation and understanding of large language models for cyber operation automation”. 2023 IEEE Conference on Communications and Network Security (CNS). IEEE, 2023.

[18] Adoma, Acheampong Francisca, Nunoo-Mensah Henry, and Wenyu Chen. “Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition”. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, 2020.

[19] Jiao, Xiaoqi, et al. “Tinybert: Distilling bert for natural language understanding”. arXiv preprint arXiv:1909.10351, 2019.

[20] Sun, Zhiqing, et al. “Mobilebert: a compact task-agnostic bert for resource-limited devices”. arXiv preprint arXiv:2004.02984, 2020.