

## **Desenvolvimento de um protótipo De software de investigação criminal utilizando processamento de linguagem natural e linguística forense.**

A Inteligência Artificial esta cada vez mais presente no dia a dia das pessoas em inúmeras aplicações. Recentemente com o aumento da criminalidade a mesma tem sido utilizada no sentido de automatizar processos desta esfera. A computação forense pode facilitar e auxiliar em uma série desses processos. A construção de novas base de dados e ferramentas que permitam o uso da inteligência artificial na computação forense é o principal enfoque desse trabalho.

O projeto apresenta uma proposta onde iremos criar uma inteligência artificial que trabalhando com linguística forense seja capaz de analisar perfis linguísticos, verificando texto e fala, visando identificar a autoria do item analisado. Especificamente almeja-se o desenvolvimento de um mecanismo com habilidade de executar análises em amostras de linguagem humana, e, que utilizando processamento de linguagem natural, seja capaz de averiguar elementos textuais, possibilitando traçar e identificar determinado perfil linguístico.

Valendo-se da Semiótica Greimasiana através do plano do conteúdo, pretendesse criar métodos a serem atribuídos ao protótipo, para que tenha eficácia no contexto investigativo na linguística forense, é um ramo da linguística, aplicada ao contexto investigativo, lidando com chamadas de emergência, comunicações de ameaça, cartas anônimas e etc.

A metodologia do projeto consiste em realizar experimentos práticos com o intuito de, através de um sistema de programação cognitiva analisar textos e por meio de padrões de uso da linguagem, identificar o perfil linguístico apontando assim, prováveis características linguísticas do locutor.

Para tal, serão utilizadas amostras de voz, formando assim o corpus linguísticos que será explorado e utilizado para comparação e identificação de padrões através de recursos baseados em inteligência artificial.

E Optou-se pela montagem de uma base de textos extraídos do twitter através de API própria. Utilizando a biblioteca Tweepy pertencente a linguagem Python, importamos publicações de pessoas públicas com o intuito de averiguar se foi de fato o próprio usuário a fazer o post ou se foi algo gerado pela comitiva de imprensa por exemplo, a ideia é aprender a maneira como o usuário analisado constrói suas publicações, pegando seus vícios de escrita e então montando um perfilhamento linguístico.

Ao final do projeto, espera-se desenvolver um sistema computacional que esteja apto a realizar perfilhamento linguístico e atribuir autoria do texto.