

# Prompt Kit: 생성형 AI 기반 탐지룰 작성 템플릿

이 Prompt Kit은 실무 환경에서 생성형 AI를 활용해 **악성코드 분석·웹공격 분석·행위 기반 탐지**를 생성까지 빠르게 수행할 수 있도록 설계된 템플릿입니다. 각 프롬프트는 “기본 지시”와 “정밀도 강화 지시”로 구성되어 있어, 본인의 숙련도나 분석 목적에 따라 필요한 부분만 적용할 수 있습니다.

- 실습 전, 반드시 **내부 도메인·경로·계정명·IP** 등 민감한 정보를 제거(**sanitize**)한 뒤 AI에게 입력해야 합니다.
- 모든 템플릿은 YARA, Snort/Suricata, SIEM Detection Rule 등 다양한 탐지 포맷에 공통적으로 적용 가능하며, 실제 운영환경에서 **오탐을 최소화하고 탐지 품질을 높이는** 방향으로 설계되어 있습니다.
- AI가 제공하는 정보는 오류가 포함될 수 있으니, 제공된 내용은 반드시 직접 검토 후 판단하시기 바랍니다.

## ◆ 1. 악성코드 분석 → YARA 룰 생성 Prompt

아래 코드를 기반으로 악성 행위를 분석해줘:

- 1) 주요 기능 요약
- 2) IOC 추출 (도메인, URL, 파일경로, 프로세스, 명령어)
- 3) 실제 탐지 가능한 특징만 정리 (난독화된 base64 등 불필요 특징 제외)
- 4) 아래 포맷으로 YARA 룰 초안 생성:
  - meta: description, author, date
  - strings: 의미 있는 문자열 최소 3개 이상
  - condition: 파일 크기 제한 + all of them
- 5) 오탐/과탐 가능성 지적
- 6) 오탐을 줄인 개선 룰 제시

### 💡 정밀도를 높일 수 있는 지시문 예시

- ✓ 문자열을 최소 3개 이상 → “행위 기반 의미가 있는 문자열만”
- ✓ code flow 또는 decode 루틴의 ‘핵심 부분’만 강조
- ✓ 일반 프로그램에서도 나타날 수 있는 문자열은 제외
- ✓ signature fragment는 5바이트 이상만 포함
- ✓ base64나 난독화 문자열은 ‘의미 기반’으로 변환 후 특징 추출
- ✓ 룰이 실제로 실행될 때 발생할 오탐 시나리오를 2개 제시

## ◆ 2. CVE 기반 → Snort/Suricata 룰 생성 Prompt

아래 HTTP 요청 또는 패킷에서 탐지 가능한 패턴을 정리해줘:

- 1) 공격 유형 설명
- 2) 탐지 포인트:
  - URI 특징
  - 헤더 기반 특징
  - 인코딩/우회 패턴
- 3) Snort 룰 생성:
  - msg, sid, rev 포함
  - content 최소 2개 이상
  - http\_uri / http\_header 구분
- 4) 오탐/과탐 가능성 지적
- 5) 오탐을 줄인 개선 룰 제시

### 💡 정밀도를 높일 수 있는 지시문 예시

- ✓ content는 최소 5바이트 이상의 희귀 문자열만 포함
- ✓ 경로 탐지와 header 탐지를 분리한 2개 룰 제공
- ✓ 공격자가 우회할 수 있는 인코딩 패턴 2개 제시
- ✓ 해당 룰을 실제 운영환경에서 튜닝하는 가이드 포함

## ◆ 3. 로그 기반 행위 분석 → 탐지를 생성 Prompt

아래 웹/시스템 로그를 기반으로 공격 시나리오를 분석해줘:

- 1) 공격 시나리오 3단계로 요약
- 2) 공격자가 수행한 행동 설명
- 3) 탐지 가능한 필드(URL, 상태코드, 파일명, 쿼리스트링 등)
- 4) Detection Rule 생성:
  - 필요 조건(AND)
  - 선택 조건(OR)
  - 시간 기반 조건 (선택)
- 5) 오탐/과탐 가능성 지적
- 6) 오탐을 줄인 개선 룰 제시

### 💡 정밀도를 높일 수 있는 지시문 예시

- ✓ 필드 중요도(High/Medium/Low) 표시
- ✓ 실제 SIEM에서 필드가 누락될 경우 대체 탐지 포인트 제안
- ✓ 오탐을 일으킬 수 있는 정상 행위 예시 2개 제시
- ✓ 보완된 룰 2차 버전 생성

---

## 4. Prompt Hygiene 체크리스트

- ☐ 내부 경로 제거
- ☐ 내부 도메인, 내부 서버명 제거
- ☐ 내부 IP → 사설IP/예시용 값으로 변경
- ☐ 계정명/사번 → generic 값으로 변경
- ☐ 민감정보가 남아있지 않은지 스스로 점검 후 입력