개인정보보호 및 정보보안 바이블

New Normal Edition

1. 개인정보 관련 용어 정의

개인정보처리자

- 업무를 목적으로 개인정보파일을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리 하는 공공기관, 법인, 단체 및 개인 등

개인정보취급자

- 개인정보처리자의 지휘 · 감독을 받아 개인정보를 처리하는 임직원, 파견근로자, 시간제근로자 등

정보주체

- 처리되는 정보에 의하여 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람
- 고객과 비(非)고객(임직원, 주주, 기자, 협력업체 직원 등)을 포함

2. Privacy By Design(PBD)

7가지 기본 원칙

번호	원칙
1	사후조치가 아닌 사전예방(Proactive not Reactive – Preventative not remedial)
2	개인정보 보호의 기본 설정(Privacy as the Default setting)
3	개인정보 보호의 내재화(Embedded Privacy into Design)
4	개인정보 보호와 서비스 목표를 둘 다 추구(Full Functionality : positive-sum, not zero-sum)
5	개인정보 생애주기 전체 보호(End-to-End Security)
6	가시성 및 투명성 유지(Visibility and Transparency – keep it open)
7	이용자 개인정보보호 존중 – 이용자 중심주의(Respect for User Privacy - keep it user centric) ▪ 이용자 중심주의를 통해 이용자의 이익을 최우선 고려 – 개인정보보호 기본 설정, 적절한 알림, 이용자 친화적인 옵션 제공 등

3. 개인정보

개인정보의 정의

제2조

“개인정보”란 살아 있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.

가목의 정보
(식별정보)

성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보

나목의 정보
(식별가능정보)

해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보

다목의 정보
(가명정보)

가목 또는 나목을 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보

4. 쉬운 결합과 어려운 결합

쉽게 결합하여 vs. 어렵게 결합하여

- 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보
- 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성, 결합 가능성 을 포함해 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려해야 함

예시

- 해킹을 통해서 얻는 정보
  - 불법 행위를 통해 얻어야 하는 정보는 입수 가능성에 포함 안됨
- 검색을 통해서 얻을 수 있는 정보
  - 이름 + 이메일 주소
  - 이름 + 휴대폰 번호
  - 입수 가능성, 결합 가능성 높음

5. 가명정보와 익명정보

가명처리

- 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것
- 가명정보는 가명처리의 결과

가명정보

- 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보

익명정보

- 시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보

6. 개인정보 수집 및 이용 동의

구분	수집 · 이용(제15조)
동의 받아 처리	1. 정보주체의 명시적 동의
동의 없이 처리	2. 법률의 특별한 규정이나 법령상 의무 준수 3. 법령 등에서 정한 소관 업무 수행(공공기관) 4. 정보주체와의 계약 체결 및 이행에 불가피하게 필요 5. 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산상 이익에 필요 6. 개인정보처리자의 정당한 이익 달성 7. 공중위생 등 공공의 안전과 안녕

6. 개인정보 수집 및 이용 동의

명시적 동의

- 법적 고지사항을 정보주체가 알기 쉽도록 알리고, 의사를 명확히 표시할 수 있는 방법으로 동의를 받아야 함

구분	수집 · 이용 시 알릴 사항(제15조)
수집 · 이용 시	1. 개인정보의 수집 · 이용 목적 2. 수집하려는 개인정보의 항목 3. 보유 및 이용 기간 4. 동의를 거부할 권리가 있다는 사실과 동의 거부에 따른 불이익이 있는 경우에는 그 불이익의 내용

6. 개인정보 수집 및 이용 동의

동의 기반 개인정보 처리의 문제점

개인정보처리자	정보주체
▪ 서비스를 제공하는 데 반드시 필요한 개인정보도 정보 주체의 동의를 받아야 함 ▪ 이용자 가입 단계가 복잡해지고, 사용성이 떨어짐	▪ 서비스를 이용하려면 어차피 동의해야 하는 항목이므로, 수집・이용 동의서 또는 개인정보 처리방침을 잘 읽지 않은 채 동의함 ▪ 개인정보 수집에 동의를 하면 이후 문제에 대해 문제 제기를 하기 어려워짐

6. 개인정보 수집 및 이용 동의

정보주체 동의 없는 개인정보 수집 예시

- 정보주체와의 계약 체결 및 체결된 계약 이행에 필요(약관도 계약의 일종)

계약 체결 예시	계약 이행 사례
▪ 보험회사가 계약체결을 위해 청약자의 자동차 사고 이력, 다른 유사보험의 가입여부 등에 관한 정보를 수집하는 경우 ▪ 회사가 취업지원자와 근로계약 체결 전에 지원자의 이력서, 졸업증명서, 성적증명서 등 정보를 수집・이용하는 경우	▪ 뉴스레터 서비스를 하는 데 필요한 이메일 주소를 수집하는 경우 ▪ 고객이 주문한 상품을 배송하기 위하여 주소, 연락처 정보를 수집하는 경우 ▪ 경품행사 시 당첨자에게 경품을 발송하기 위해 주소와 연락처 정보를 수집하는 경우 ▪ 쇼핑몰이 주문 시 포인트를 지급하기로 약정하고 주문정보를 수집하는 경우

7. 민감정보, 고유식별정보, 주민등록번호

민감정보

- 민감정보의 범위(개인정보보호법 제23조)
  - ① 사상 · 신념
  - ② 노동조합 · 정당의 가입 · 탈퇴
  - ③ 정치적 견해
  - ④ 건강, 성생활 등에 관한 정보
  - ⑤ 유정정보
  - ⑥ 범죄경력 정보
  - ⑦ 생체인식 특징 정보
  - ⑧ 인종 · 민족에 관한 정보

7. 민감정보, 고유식별정보, 주민등록번호

고유식별정보

- 고유식별정보의 범위(개인정보보호법 제24조, 제24조의2)
  - 법령에 따라 개인을 고유하게 구별하기 위하여 부여된 식별정보로서 대통령령(개인정보보호법)으로 정하는 정보
    ① 주민등록번호
    ② 여권번호
    ③ 운전면허번호
    ④ 외국인등록번호
  - 주민등록번호는 고유식별정보 중 특별한 지위에 있음

7. 민감정보, 고유식별정보, 주민등록번호

민감정보 · 고유식별정보의 처리

- 원칙적으로 처리 금지
- 처리할 수 있는 경우
  - 정보주체의 별도 동의를 받은 경우
    - ✓ 별도 : 다른 동의와 구분(분리)하여
  - 명시적 동의 : 정보주체에게서 동의를 받을 때 사전 알림 사항
  - 법령에서 민감정보 · 고유식별정보의 처리를 허용하는 경우 (정보주체의 동의 없이 처리)
  - 법령에서 명시적으로 민감정보 · 고유식별정보 (또는 그것의 종류)를 열거하고 그것의 처리를 허용
  - 법령 : 법률, 시행령, 시행규칙

8. 업무를 위한 영상 촬영

이동형 영상정보처리기기의 사용

원칙	업무를 목적으로 공개된 장소에서 이동형 영상정보처리기기로 촬영은 일반적으로 금지 (개인정보보호법 제25조의2)
예외	제15조(개인정보의 수집 · 이용) 제1항 각 호(제1호~제7호)의 어느 하나에 해당하는 경우

- 촬영 사실을 명확히 표시하여 정보주체가 촬영 사실을 알 수 있도록 하였음에도 불구하고 촬영 거부 의사를 밝히지 아니한 경우
- 이 경우 정보주체의 권리를 부당하게 침해할 우려가 없고 합리적인 범위를 초과하지 아니하는 경우로 한정

9. 정보주체의 권리 보장

권리 보장 사례

본인이 찍힌 CCTV 영상을 요청했는데, 시간만 끌고 보여주지 않아요

건물관리・경비・청소 등 시설유지관리 용역을 제공하는 업체가 CCTV 설치 업체와
설치 계약을 체결하고 신고인이 입주한 건물의 관리업무를 위탁・수행하며 건물 내 CCTV 설치・운영

- CCTV 설치 업체가 신고인의 요청한 본인이 촬영된 특정일의 CCTV 영상에 대한 열람 요구를 위탁사 승인이 필요하다는 이유로 거부
- 신고인이 경찰 입회하에 열람을 요구하였으나 다른 사람의 영상이 포함되어 있다는 이유로 열람 거부
- 신고인이 피심인(건물관리인)에게 내용증명을 발송했으나 업체는 이에 대응하지 않음
- 피심인은 신고인이 요구한 CCTV 영상에 타인이 촬영되어 있어 바로 조치하지 못하고 사본 제공

9. 정보주체의 권리 보장

열람요구권(제35조)

- 정보주체(또는 14세 미만 아동의 법정대리인)가 자신의 개인정보에 대한 열람을 요구할 수 있는 권리
- 개인정보처리자는10일 이내에 열람하도록 하거나, 정당한 거절 사유가 있으면, 그 사유를 알려야 함

정정 · 삭제권(제36조)

- 정보주체(또는 14세 미만 아동의 법정대리인)가 자신의 개인정보에 대한 정정 · 삭제를 요구할 수 있는 권리
- 개인정보처리자는 10일 이내에 처리하고 결과를 알려야 함
- 정당한 거절 사유가 있으면, 그 사유를 알려야 함

열람 제한 · 거절 사유

- 법률에 따라 열람이 금지되거나 제한되는 경우
- 다른 사람의 생명 · 신체를 해할 우려가 있거나, 다른 사람의 재산과 그 밖의 이익을 부당하게 침해할 우려가 있는 경우
- 공공기관의 해당 업무를 수행에 중대한 지장을 초래하는 경우(조세, 성적 평가, 시험 및 자격 심사 등)

정정 · 삭제 거절 사유

- 다른 법령에서 그 개인정보가 수집 대상으로 명시되어 있는 경우
- 14세 미만 아동의 법정대리인도 동일한 권리를 행사할 수 있음

9. 정보주체의 권리 보장

처리정지권(제37조)

- 정보주체(또는 14세 미만 아동의 법정대리인)가 자신의 개인정보에 대한 처리 정지를 요구할 수 있는 권리
- 개인정보처리자는 10일 이내 처리 정지, 파기하여야 함
- 정당한 거절 사유가 있으면, 그 사유를 알려야 함

동의철회권(제37조)

- 이용자가(또는 14세 미만 아동의 법정대리인)가 자신의 개인정보에 대한 동의 철회를 요구할 수 있는 권리
- 개인정보처리자는 10일 이내 동의 철회하고, 파기해야 함
- 정당한 파기 거절 사유가 있으면 그 사유를 알려야 함

처리정지 거절 사유

- 법률에 특별한 규정이 있거나, 법령상 의무를 준수하기 위하여 불가피한 경우
- 다른 사람의 생명・신체를 해할 우려가 있거나 다른 사람의 재산과 그 밖의 이익을 부당하게 침해할 우려가 있는 경우
- 공공기관이 해당 개인정보를 처리하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우

동의 철회는 거절 불가

- 파기 거절 사유 = 처리정지 거절 사유

9. 정보주체의 권리 보장

사업자(개인정보처리자)의 의무

- 타인의 개인영상정보가 포함되어 있으면 열람 거절 사유는 됨
- 다만 10일 이내에 최소한 거절 사유를 알려야 함
- 타인의 영상정보를 모자이크, 해당 타인의 동의를 얻는 등의 방식으로 제공해야 함

- 사업자는 정보주체가 자신의 열람, 정정삭제, 처리정지, 동의철회 요구권을 행사할 때를 대비하여
- 정보주체가 해당 권리를 행사할 수 있는 창구를 개설하고, 처리 조직과 절차를 준비하여 10일 이내에 대응할 수 있어야 함
- 창구 : 개인정보 처리방침에 개인정보보호책임자 이메일 등을 이용할 수 있음

9. 정보주체의 권리 보장

자동화 결정에 대한 정보주체의 권리(2024.3.15 시행)

- 자동화된 결정
  - AI 등을 적용한 완전히 자동화된 시스템으로 개인정보를 처리하여 이뤄지는 결정
  - AI 기반 면접, 자기소개서 평가, 신용평가
- 정보주체의 권리
  - 결정 거부권 결정을 거부할 수 있는 권리
  - 예외 : 제15조(개인정보의 수집 · 이용) 제1항 제1호 · 제2호 · 제4호에 따라 이루어지는 경우
  - 설명 요구권 결정에 대한 설명 요구

제15조(개인정보의 수집 · 이용) 제1항

- 정보주체의 동의를 받은 경우
- 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
- 정보주체와 체결한 계약을 이행하거나 계약을 체결하는 과정에서 정보주체의 요청에 따른 조치를 이행하기 위하여 필요한 경우

10. 개인정보 유출 대응

개인정보 유출 등의 통지

분류	내용
통지 요건	▪ (1명 이상의 정보주체의) 개인정보가 유출 등이 되었음을 알게 되었을 때 ▪ 유출 등 : 분실 · 도난 · 유출
통지 기한	▪ 유출사실을 알았을 때 지체없이 – 72시간 이내(영 제39조 제1항) ▪ [예외] 72시간이 넘어 통지할 수 있는 사유 – 사유 해소되면 지체없이 통지 ▪ 유출 등이 된 개인정보 의 확산 및 추가 유출 등을 방지하기 위하여 접속경로의 차단, 취약점 점검 · 보완, 유출 등이 된 개인정보의 회수 · 삭제 등 긴급한 조치가 필요한 경우 ▪ 천재지변이나 그 밖에 부득이한 사유로 인하여 72시간 이내에 통지하기 곤란한 경우
통지 방법	▪ 개별 통지(서면 등의 방법) : 서면, 전자우편, 팩스, 전화, 문자 또는 이에 상당하는 방법 ▪ [예외] 연락처가 없으면 30일 이상 인터넷 홈페이지에 게시 (인터넷 홈페이지가 없으면, 사업장 등 보기 쉬운 장소에 30일 이상 게시)
적용 조항	▪ 법 제34조(개인정보 유출 등의 통지 · 신고), 시행령 제39조(개인정보 유출 통지의 방법 및 절차)

10. 개인정보 유출 대응

개인정보 유출 등의 통지

분류	내용
통지 내용	■ 다음 5가지를 정보주체에 알려야 함 ① 유출 등이 된 개인정보의 항목 ② 유출 등이 된 시점과 그 경위 ③ 유출 등으로 인하여 발생할 수 있는 피해를 최소화하기 위하여 정보주체가 할 수 있는 방법 등에 관한 정보 ④ 개인정보처리자의 대응조치 및 피해 구제절차 ⑤ 정보주체에게 피해가 발생한 경우 신고 등을 접수할 수 있는 담당부서 및 연락처
	■ [통지 내용 예외] 위 ① 또는 ②를 파악하지 못한 경우, 파악한 것과 ③~⑤를 먼저 정보주체에 통지(우선 통지)하고, 나머지는 파악한 즉시 통지(추후 통지)

10. 개인정보 유출 대응

개인정보 유출 등의 신고

분류	내용
신고 요건	■ 개인정보 유출 등이 되었음을 알게 되었을 때 다음 중 하나에 해당하는 경우(영 제40조 제1항) ① 1천 명 이상의 정보주체에 관한 개인정보가 유출 등이 된 경우 ② 민감정보, 고유식별정보가 유출 등이 된 경우 ③ 외부로부터의 불법적인 접근에 의해 개인정보가 유출 등이 된 경우 해킹에 의한 개인정보 유출 ■ [예외] 신고하지 않을 수 있는 사유 ■ 개인정보 유출 등의 경로가 확인되어 해당 개인정보를 회수·삭제하는 등의 조치를 통해 정보주체의 권익 침해 가능성이 현저히 낮은 경우
신고 기한	■ 유출사실을 알았을 때 지체 없이 – 72시간 이내(영 제40조 제1항) ■ [예외] 72시간 넘어 신고할 수 있는 사유 ■ 천재지변이나 그 밖에 부득이한 사유로 인하여 72시간 이내에 신고하기 곤란한 경우

10. 개인정보 유출 대응

개인정보 유출 등의 신고

분류	내용
신고 내용	■ 통지 내용과 동일, 신고 내용 예외도 통지 내용 예외와 동일
신고처	■ 개인정보보호위원회 또는 한국인터넷진흥원
적용 조항	■ 법 제34조(개인정보 유출 등의 통지・신고), 시행령 제40조(개인정보 유출 등의 신고)

10. 개인정보 유출 대응

특수한 상황의 통지와 신고

- 가명정보 유출 시
  - 신고 의무는 있지만, 통지 의무는 없음
- 수탁자의 개인정보 유출 시
  - 위탁자(개인정보처리자)에 통지・신고 의무가 있음
  - 수탁자에도 통지・신고의 의무가 있음(제26조 제8항에서 제34조 적용 적시)

11. 개인정보 위기관리

개인정보 위기 관리의 목적

개인정보 이슈의 예방

개인정보 사건 · 사고가 위기로 확대되는 것 방지

위기대응 및 피해 복구

위기 예방

피해 최소화

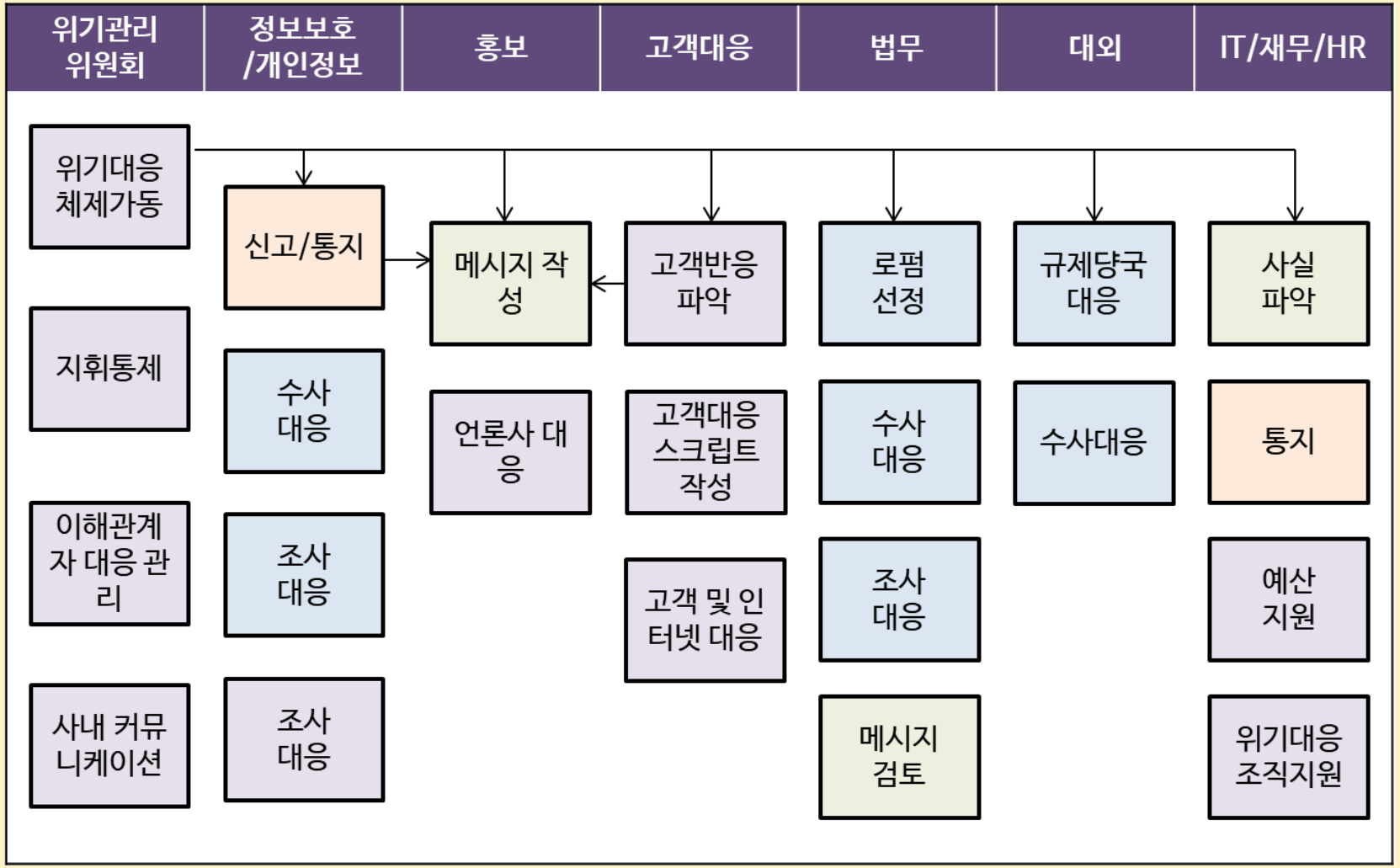
11. 개인정보 위기관리

개인정보 위기 관리의 목적

구분	단계	업무
위기 전 업무	위기 예방	▪ 개인정보의 기술적 · 관리적 보호조치 ▪ 개인정보 이슈 탐지와 조기 대응 ▪ 일상적 위기관리
	위기 대비	▪ 위기관리 체계의 수립, 관리 ▪ 위기대응 모의 훈련
위기 후 업무	위기 대응	▪ 법규 준수 ▪ 이해관계자 대응
	복구	▪ 데이터 · 서비스 · 시스템의 복구 ▪ 고객 신뢰도, 브랜드 가치의 회복

11. 개인정보 위기관리

위기 관리 조직과 대응 프로세스



11. 개인정보 위기관리

모의훈련

예방, 대비	피싱메일(탐지 및 대응)
대응	통지문, 신고문 쓰기, 언론 대응
복구	프로그램 소스 복구, 데이터 복구고객 신뢰 회복

12. DoS 공격과 DDos 공격

DoS(Denial of Service) 공격	DDoS(Distributed Denial of Service) 공격
▪ (직역) 서비스 거부 공격 ▪ 서비스 제공자가 서비스를 제대로 제공하지 못하도록 하는 공격 ▪ (의미) 서비스 중단(마비) 공격 방법 ▪ 스마트폰의 배터리를 제거 ▪ PC의 전원 끄기 ▪ 웹 서버에 대규모 인터넷 데이터(트래픽) 송신	▪ (번역) 분산 서비스 거부 공격 ▪ 여러 PC나 전자기기에서 표적에 대해 대규모 데이터를 송신함으로써 서비스가 중단・지연되도록 하는 공격 방법 ▪ F5 공격(http) ▪ 많은 PC나 공유기 등에 악성코드를 설치 (좀비 기기의 확보)해 표적에 대규모 트래픽을 송신

12. DoS 공격과 DDos 공격

DDos 공격에 대한 대책

DDos 대응 솔루션은 공격 트래픽은 차단하고 정상 트래픽은 통과시켜 정상적인 서비스가 이뤄지도록 함

기업	개인
▪ DDos 대응 장비 구축 ▪ 클라우드서비스, 인터넷회선서비스 등에서 제공하는 DDos 대응 서비스 활용	▪ 내가 사용하는 IT 기기가 좀비가 되지 않도록 유의 • PC가 악성코드에 감염되지 않도록 주의 • 컴퓨터 백신 설치, 운영체제 자동 업데이트 등 • 인터넷 공유기, IP카메라 등의 초기 비밀번호 변경

13. 랜섬웨어

랜섬웨어 대응책

구분	내용
사업 연속성(BCP)	■ 사업 중단 여부와 기간은 어떻게 준비하느냐에 달려 있어요. 사업 책임자와 CFO 등 고위 임원, 관련 부서가 주도적으로 준비 ■ 사업 지속하기 필요한 시스템과 데이터, 개발 소스, 연구 문서 등 선정, 인력과 조직, 예산 투자 필요
악성코드 대응	■ 보안 프로그램(안티바이러스 등) 설치 • 실시간 감시, 자동 업데이트, 주기적 전체 검사, 운영체제 자동 업데이트, 표적 공격(APT) 대응, 피싱 메일 대응 ■ 개인이 웹하드, P2P 등 악성코드가 많이 배포되는 곳 방문하거나 다운받지 않기, 피싱 메일 유의 • 꼭 필요한 메일만 열어 보기, 첨부파일 다운받지 않기
백업 및 복구 (훈련)	■ 안전한 곳에 백업 : 오프라인이나 랜섬웨어가 접근할 수 없게 통제된 상태로 보관 ■ 복구 훈련 : 신속하고 정확하게 복구 가능

14. 사이버 전쟁

러시아 - 우크라이나 전쟁

- 2022년 2월 24일, 러시아가 우크라이나를 침공하면서 시작한 러시아-우크라이나 전쟁은 물리 공간뿐 아니라 사이버 공간에서도 전투가 이뤄지는 하이브리드 전쟁
- 양 국가의 정규군뿐 아니라 친러, 친우크라 해커그룹들이 사이버 공간에 상대 국가나 상대를 지원하는 국가, 관련 기업까지 무차별 공격함으로써 민간기업에까지 피해 확산

미국 - 중국 전략 전쟁

- 세계의 두 강대국인 미국과 중국은 정치, 군사, 경제 등 다양한 분야에서 상대방의 고급 정보를 빼내기 위한 치열한 사이버 전투를 벌이고 있음

15. 보안 취약점

임직원의 보안 행동에 긍정적 영향을 미치는 요인

규정 준수가 IT시스템 보안에 도움이 됨

규정 위반 시 탐지될 것임

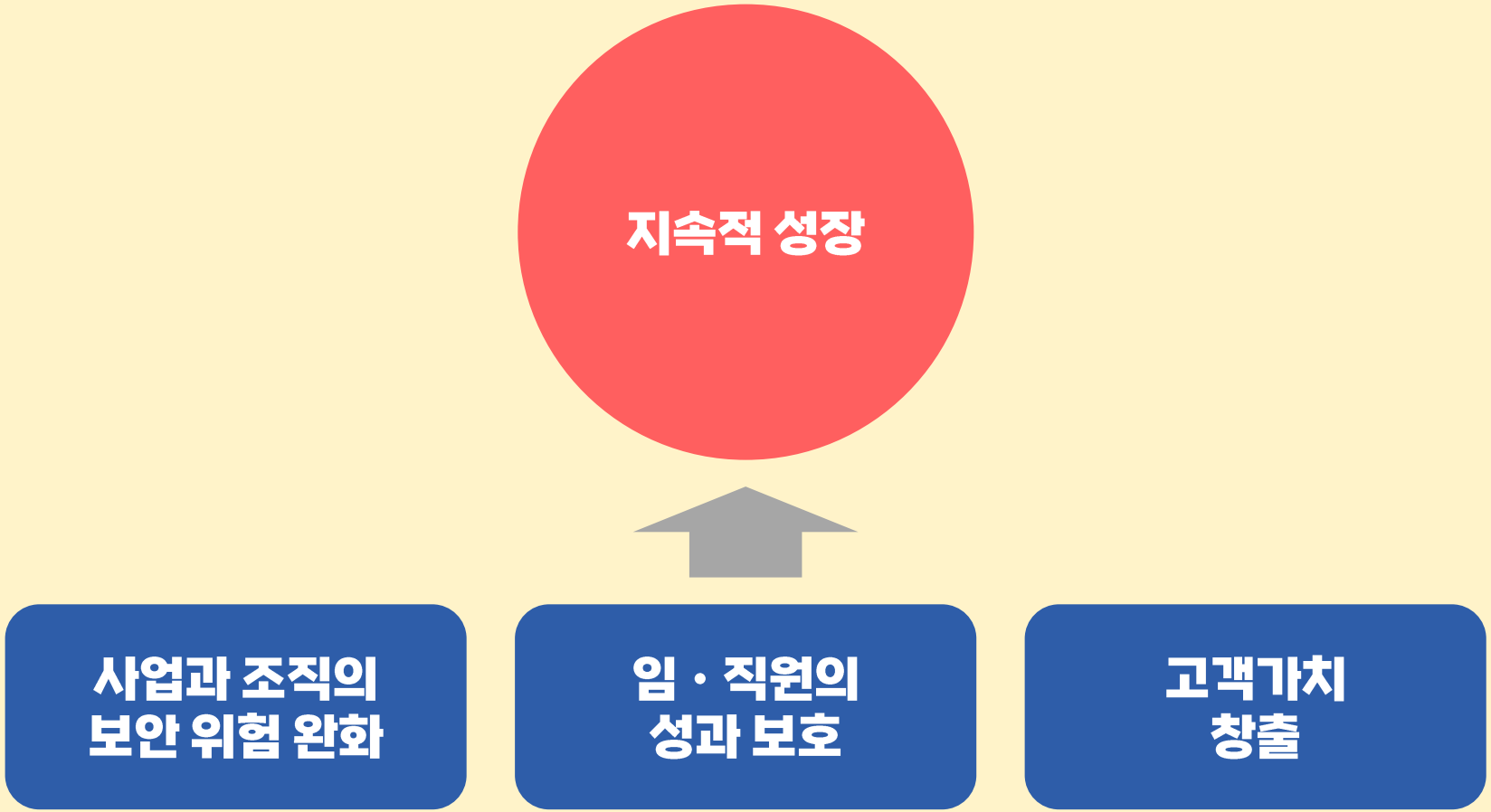
자신이 규정을 준수할 거라는 상사 또는 동료의 믿음

동료들이 규정을 잘 준수함

보안 규정 준수

15. 보안 취약점

정보보안의 목적 – 사업적 측면



16. 인공지능의 문제점과 악용

ChatGPT의 문제점

1 문맥 이해의 한계

- ChatGPT는 입력 문맥을 이해하는 데 한계가 있음
- 긴 문장이나 복잡한 문맥에서 이전에 언급된 정보를 정확하게 기억하는 것이 어려울 수 있음

2 긴 문장 처리의 어려움

- 모델은 제한된 문맥 크기로 인해 긴 문장을 처리하는 데 어려움을 겪을 수 있음

16. 인공지능의 문제점과 악용

ChatGPT의 문제점

3 정확성 보장의 어려움

- 모델은 학습 데이터를 기반으로 생성되기 때문에, 그 데이터에 있는 오류, 편향, 부정확한 정보 등을 반영할 수 있음
- 따라서 모델이 항상 정확하고 신뢰할 수 있는 정보를 생성하는 것은 아님

4 데이터 편향과 불순화

- 학습 데이터에 포함된 편향된 정보나 불순화된 언어 패턴은 모델의 생성물에 영향을 미칠 수 있음
- 이로 인해 모델이 특정 그룹이나 주제에 대해 편향될 수 있음

16. 인공지능의 문제점과 악용

ChatGPT의 문제점

5 감정적 민감성과 부적절한 응답

- 모델은 감정을 이해하려고 노력하지만, 때로는 부적절한 감정적 응답이나 민감한 주제에 대한 부적절한 답변을 생성할 수 있음

6 제한된 상식과 현실 세계 이해의 한계

- 모델은 학습 데이터에서 얻은 상식을 기반으로 함
- 실제 세계의 최신 정보나 특정 도메인의 전문 지식이 부족할 수 있음

16. 인공지능의 문제점과 악용

ChatGPT의 문제점

7 대화 지속 능력의 한계

- 모델은 일정한 대화 주기 이상으로 진행되면 무의미한 또는 반복적인 응답을 생성할 수 있음

8 인공지능 공격에 취약

- 악의적인 사용자가 모델을 사용하여 유해한 내용을 생성하거나 다양한 종류의 사회 공학 공격을 시도할 수 있음

16. 인공지능의 문제점과 악용

ChatGPT의 악용

중요 정보 추출	피싱 메일, 피싱 사이트 생성	악성코드 생성
▪ ChatGPT와의 대화를 통해 ChatGPT 학습했던 기업의 중요 데이터나 개인정보를 추출	▪ 피싱(Phishing) 목적이라는 걸 밝히지 않고, 환경과 요건을 규정하여 메일이나 웹 사이트 생성을 요구하면 그 요건을 충족하는 피싱 메일과 피싱 사이트(소스)를 생성해줌	▪ 환경과 요건을 규정하여 악성행위를 하는 코드 생성을 요구하면 그 요건을 충족하는 코드를 생성해 줌

17. AI에 대한 공격

학습 단계 공격

- 데이터 오염 공격
  - 인공지능이 학습하는 데이터를 여러 방법을 통해 오염시켜 잘못된 결과를 생성하거나 성능을 떨어뜨리는 공격 방법
- 회피 공격
  - 노이즈를 추가한 적대적 데이터를 입력하여, AI 모델이 잘못된 판단을 하도록 유도하는 공격
  - 판다 사진에 노이즈를 추가함으로써 AI가 긴팔원숭이로 잘못 인식하게 할 수 있음
  - 자율주행차에 탑재된 AI시스템이 교통표지판이나 사람을 잘못 인식하면 큰 문제가 발생할 수 있음

18. OWASP Top 10 for LLM

OWASP Top 10

Open Web Application Security Project라는 단체에서 발표한 분야별 상위 취약점 10개

- 프롬프트 주입
- 데이터 유출
- 부적절한 샌드박싱
- 무단 코드 실행
- 서버 측 요청 위조
- 대규모언어모델(LLM) 생성 콘텐츠에 대한 과도한 의존
- 부적절한 AI 정렬
- 불충분한 접근 통제 조치
- 부적절한 오류 처리
- 학습 데이터 중독(오염)

19. 보안에서 AI 이용하기

보안관제	보안 위협의 탐지 및 대응의 자동화
클라우드	클라우드에 쌓이는 각종 로그 분석을 통한 이상행위 탐지
소프트웨어	소프트웨어에 존재하는 보안취약점 탐지
악성코드	다양한 실행파일 중 악성파일 탐지, 알려지지 않은 악성코드 탐지 가능

20. 인공지능 규제 동향

개인정보보호법 제37조의2(자동화된 결정에 대한 정보주체의 권리 등)

- 자동화된 결정
  - AI 등을 적용한 완전히 자동화된 시스템으로 개인정보를 처리하여 이뤄지는 결정
  - AI 기반 면접 또는 자기소개서 평가, 신용평가
- 정보주체의 권리
  - 결정 거부권
    - ✓ AI에 의한 결정을 거부할 수 있는 권리
    - ✓ (기업) 자동화 결정에서 제외, 정보주체가 요구하면 사람이 개입하여 재처리
  - 설명 요구권
    - ✓ 결정에 대한 설명을 요구할 수 있는 권리
    - ✓ (기업) 이해하기 쉽게 설명

20. 인공지능 규제 동향

유럽연합 인공지능법(AI Act)

정의(제3조)

다양한 수준의 자율성을 가지고 작동하도록 설계되어 명시적 또는 묵시적 목표에 따라 물리적 또는 가상 환경에 영향을 미치는 예측이나 추천, 결정과 같은 결과물을 생성할 수 있는 기계 기반 시스템을 의미

금지된 관행(제5조)

- 알려진 또는 예측된 성격 특성, 연령, 신체적 또는 정신적 장애, 사회적 및/또는 경제적 지위를 악용하는 행위 (치료 목적 제외)
- 민감하거나 보호되는 속성에 따라 사람을 분류하는 생체인식분류시스템의 출시・서비스・사용 (치료 목적 제외)
- 공공장소에서의 실시간 원격 생체 인식 시스템 사용
- 개인적 또는 성격적 특성을 기반으로 개인 또는 집단의 사회적 점수를 평가 또는 분류
- 인터넷 또는 CCTV 영상에서 얼굴 이미지 스크래핑을 통해 얼굴 인식 데이터베이스를 생성, 확장하는 행위

20. 인공지능 규제 동향

고위험 AI 시스템

- AI시스템이 제품의 안전 요소이거나 안전에 관한 제품인 경우
- 생체인식시스템과 생체 기반 시스템
- 생체식별시스템
- 감정 인식 시스템을 포함한 생체 인식 또는 생체 기반 데이터를 기반으로 사람의 개인적 특성을 추론 등