

## NEWSLETTER

Vol. 3 9 May 2018

1

## TISQA 2018 컨퍼런스 : Why testers need AI, more for QA pros

본 내용은 TechBeacon의 기사를 발췌/요약한 기사로 작성자는 TechBeacon의 Editor Mitch Pronschinske 입니다.

### ◆ 행사소개

TISQA (Triangle Information Systems Quality Association)는 미국의 소프트웨어 품질 관련 비영리 단체로 네트워킹, 교육 및 전문 개발 기회를 제공하고 있으며 정보 시스템 품질 관리 방식 및 원칙을 개선하도록 장려하고 있으며, 매년 Triangle's Software Quality Conference를 개최하고 있다. 테스트 기술, 테스트 관리 및 메트릭, 테스트 자동화, 모바일 및 성능 테스트, 일부 Devops 주제에 대한 세션을 포함한 테스트에 초점을 맞춘 컨퍼런스이다.

가장 최근에 개최된 TISQA 2018에서는 AI가 소프트웨어 테스트의 복잡성을 자동화할 수 있는 유일한 방법이라는 이유 때문에 테스트 업계의 AI 도입에 대한 다양한 논의가 있었다.

### ◆ 데이터 중심의 모바일 테스트 전략

: Jason Arbon ( Microsoft 및 Google 엔지니어 및 테스터, Test.ai. CEO겸 공동 설립자)

1	게임 앱은 좋은 게임을 만드는 것에 집중해라. 재미있지만 하면 대부분의 사람들은 안정성, 정밀성 및 보안 문제에 크게 신경 쓰지 않는다.
2	금융 앱과 같은 실용적인 앱은 가용성이 가장 큰 관심사이다. 안정성은 앱이 유용한 경우에 큰 문제가 아니다.
3	다방면에서 앱의 성능을 통제하는 것은 어렵다. 따라서 사용자에게 네트워크 연결로 인해 성능에 영향이 생기거나 통제 범위를 벗어나는 일이 발생했을 경우 알려주도록 한다.
4	앱을 개발할 때, 경쟁사의 앱 리뷰를 읽고 어떤 클레임이 있는지 확인해라. 이를 바탕으로 경쟁 업체보다 더 나은 제품을 만들도록 한다.
5	수천 개의 앱을 분석한 결과, 평균 리뷰는 앱 릴리즈와 관계가 없다. 즉, 2주에 한번씩 새 버전으로 릴리즈 할 필요가 없다.

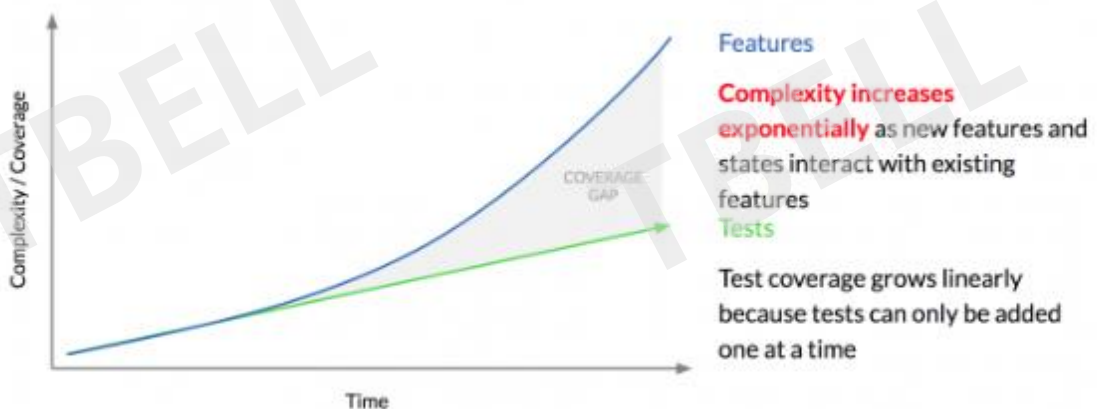
예를 들면, Arbon이 관여한 앱 중 하나는 Apple 앱 스토어 에서 몇 년간 상위 30위에 있었다. 처음에 그들은 자주 업데이트를 했지만 일정 수준의 검토가 끝나고 난 후에 자주 릴리즈 하지 않는 방향으로 노선을 바꿨다. 그 이유는 이미사람들이 좋아하는 것에 변화를 주는 것은 위험하기 때문이다.

스타벅스는 앱 업데이트 후 좋지 않은 리뷰로 별점이 떨어졌고, 이를 본 경영진은 테스트 팀 모두를 해고했다. 그 조치가 좋았는지 나쁜지는 중요하지 않다. 엔지니어와 테스터는 단순히 앱이 제대로 작동하는 지가 아닌 항상 그 이상의 것을 생각해야 한다는 것이다.

### AI 테스트가 가장 좋은 마지막 희망인가?

Arbon는 강연 후반에 현재 행해지고 있는 테스트 자동화 실습의 무의미 함에 대해 이야기 하였다. 이미 많은 테스터들이 경험해서 알고 있겠지만, 대규모 테스트 자동화 작업은 낙관적인 기대로 시작하여 절망으로 끝난다. 하지만 Arbon은 테스트 자동화는 꼭 필요하다라고 강조하였다.

테스트 자동화는 단순한 웹 어플리케이션에 비해 변수 및 환경이 많기 때문에 문제가 많이 발생한다고 한다. 때문에 클라우드 테스트 및 크로스 플랫폼 테스트 툴로는 문제를 해결하기 어렵다. 아래의 그래프에서 알 수 있듯이 앞으로 개발자가 앱에 추가하는 복잡성이 테스트 개발을 크게 능가할 것이라고 한다.



지난 몇 년간 개발은 많이 바뀌었지만 테스트 전략과 기술은 10년이 넘도록 혁신적인 발전을 하지 않았기 때문에 AI가 테스트 자동화에 혁신을 가져다 줄 거라고 한다. 앱의 복잡성이 급격히 늘어남에 따라 간단한 테스트 케이스를 빠르게 처리 해야만 한다.

그의 회사는 앱 스토어의 모든 앱을 신경망에 보내서 로그인 필드와 햄버거 메뉴와 같은 요소를 인식 할 수 있게 했으며, 앱이 복잡한 동작을 하지 않는 한 functional

test 케이스를 만들 수 있도록 하였다. 많은 기업들은 이와 같은 문제를 해결하기 위해서 노력하고 있으며 이로 인해 테스터에게 많은 영향을 주게 될 것이다.

### ◆ 어떤 테스트를 자동화해야 하는가?

: Angie Jones (Twitter 수석 소프트웨어 엔지니어)

이 주제는 테스터가 직면하는 가장 큰 문제 중 하나이다. 테스트를 자동화하여 컴퓨터가 많은 작업을 수행하도록 하는 것은 바람직하다. 하지만 모든 잠재적 버그에 대해서 테스트 자동화를 할 수는 없다. **과도한 자동화는 비용이 많이 들 수도 있기** 때문이다. 테스트가 자동화되어 있다는 것은 관리해야 할 항목이 많아지고 개발 수명 주기가 짧아진다는 것을 의미한다.

중복 자동화는 더더욱 어렵다. 실제로 50개의 오류가 없어도 50개의 테스트 오류가 발생할 수 있다. 이러한 오류는 대부분의 테스트가 소프트웨어를 통과하는 경로가 비슷하기 때문이다.

Twitter에는 이것을 전제로 **자동화 테스트 모델**을 수립하여 테스트를 실행하고 있다. 아래에 Twitter에서 자동화를 할 것인지에 대한 의사 결정을 내리는 데 사용되는 4가지 메트릭을 소개하고자 한다.

메트릭	평가 항목	
	1	2
위험	이 기능을 고객이 얼마나 자주 사용하는가?	이 기능을 사용하지 못하게 되면 고객에게 어떤 영향을 주는가?
값	이 테스트가 새로운 정보를 제공하는가? 아니면 다른 많은 테스트와 중복되는가?	버그 트래커를 사용하면 얼마나 빨리 문제를 해결 할 수 있는가?
	참고: 테스터는 상황이 보이는 것처럼 뚜렷하지 않다는 것을 깨닫지 못하는 경우가 많다. 예를 들어, 이모티콘을 사용한 트위터와 이모티콘을 비교하는 테스트는 많이 다르지 않다.	
비용 효율성	자동화된 테스트를 얼마나 빨리 스크립트로 작성할 수 있는가?	자동화된 테스트를 스크립트로 작성하는 것이 얼마나 쉬운가?
	참고 : 이 항목에서 시간이 덜 걸리는 테스트는 낮은 점수가 아니라 높은 점수를 받아야 한다.	
히스토리	과거에 이 기능의 코드 베이스 영역에서 발생한 고장은 얼마나 되는가?	과거에 이 테스트의 실패 건수는 얼마나 되는가?

테스터는 위 4개 메트릭의 두 항목에 대해서 각각 1에서 5사이의 점수를 부여한다. 따라서 점수는 1~25점 사이가 된다. 그 후 각 메트릭의 점수를 합산한다. 이렇게 도출된 점수를 통해 다음의 <점수별 자동화 도입여부>의 기준에 따라서 자동화 여부를 결정하면 된다.

<점수별 자동화 도입여부>

점수	자동화 도입여부
75 ~ 100	도입
26 ~ 74	가능하다면 도입
4 ~ 25	도입하지 않음

<예시>

메트릭	1	2	메트릭별 점수
위험	5	4	20
값	5	4	20
비효율성	5	4	20
히스토리	2	2	4
총계			64

위의 <예시>를 보면 각 메트릭의 평가항목 1, 2의 점수의 곱인 메트릭별 점수의 총계의 합이 64가 되므로 이 경우에는 ‘가능하다면 자동화를 도입’ 하는 것으로 결정을 할 수 있다.

이 점수를 통해 자동화 도입 여부뿐만이 아니라 우선 순위도 결정할 수 있는데, 점수가 높을수록 테스트에 자동화 도입을 빨리 해야 한다는 것을 의미한다.

## 2

## AI Bot이 사람에게서 QA testing 일자리를 빼앗을 것인가?

본 내용은 TechBeacon의 기사를 요약한 것으로 작성자는 Beaufort Fairmont의 CEO Paul Merrill입니다.

많은 사람들이 SW 테스트 업계에 기계학습이 도입이 될 거라고 전망하고 있다. 기계학습이 도입되어도 사람이 하는 테스트 업무가 사라지지 않는 않을 것이다. 단지 그 역할이 바뀌게 될 뿐이다.

### 우리가 알고 있는 테스트는 이제 끝인가?

Appdiff은 AI 기반 모바일 테스트 플랫폼을 구축하여 사람이 개입하지 않고도 모바일 앱 테스트를 수행 할 수 있는 모바일 앱 테스트 프로그램을 목표로 하고 있다. 이 프로그램의 개발자들에게 있어 첫번째 도전은 Bot에게 iPhone 및 Android 핸드폰으로 제스처 하는 방법을 가르치는 것이었다. 그리고 이 에이전트가 다른 앱에서 어떤 행동을 할 수 있는지, 앱을 이용해서 어떻게 탐색해야 하는지, 어떤 종류의 입력 값을 사용해야 하는지에 대한 문제들을 해결했다.

그러나 모바일 앱과 상호 작용하는 방법을 배우는 것은 테스트가 아니다. 그래서 결함을 찾기 위한 휴리스틱 (heuristics) 비즈니스 도메인에 대한 이해 능력 등을 통

해 응용프로그램에서 사용하는 퍼지 (fuzzing) 이상으로 발전 시켰다. 여기서 멈추지 않고, 기계 학습 알고리즘에 주어진 동작의 결과로 결함이 도출 될 가능성이 있는지를 알게 하였으며, 행동과 결과가 기대치에서 벗어난 것 같을 때 확실성을 알 수 있도록 하였다.

Appdiff은 모바일 앱 테스트에 있어서 약 90 %정도의 커버리지를 가지고 있다. 테스터의 경우 커버리지가 90%인 경우는 거의 없고, 나머지 10 %는 대부분의 기업들이 테스트에 투자하기에는 너무 비싸거나 너무 복잡하다. 또한 응용 프로그램과 상호 작용하여 마지막 10 %까지 테스트 할 만큼의 능력을 가진 테스터도 거의 없다. 하지만 APPdiff는 이 모든 것이 가능해질 것이라고 한다.

### 인간의 차별화 요소

PrecisionLender의 수석 분석 책임자 인 George Neal은 AI가 테스터를 대신하지 않을 것이라고 했다. 하지만 앱이 무엇을 할지 모든 경우에서 알 수 없기 때문에 기계 학습을 앱에 도입 할 때 테스트가 훨씬 더 어려워 질 것이라고 했다. 어려운 문제가 있을 때 기계학습은 확실성이 아닌 가능성에 따라 선택을 할 것이다.

"인간이 잘하는 일을 하기 싫어하는 사람들에게 미래는 매우 무서운 곳."이라고 Neal은 말한다. 아이디어란 인간이 창의력, 탐구, 이해, 분석 및 지식을 적용하는 것이다. 따라서 이러한 활동을 좋아하지 않는 사람들은 장래에 그들이 좋아하는 일을 하는 데 어려움을 겪을 것이라는 것이다.

Neal에게 기계 학습이 테스트를 대신 할 것인지 물어봤을 때 대답하지 않았으나, 그는 "점점 더 테스트가 어려워 질 것"이라고 덧붙였다.

지금까지 우리는 소프트웨어 개발 수명 주기 전반에 걸쳐 테스트의 원칙을 결정론적 활동으로 진행해 왔다. 컴퓨터는 우리가 옳고 그름을 확신하며 결정할 수 있는 결과만을 만들어 낼 것이다. 하지만 기계 학습은 더 크고 복잡한 문제에 비 결정적 결과를 도출 할 수 있을 것이다. 과거에 테스터로서 가장 어려운 것은 비 결정적인 것을 찾아내는 것으로 예를 들면, 멀티 스레드 환경에서 결함을 찾는 데 필요한 전체 조건을 정의하는 것과 같은 것이다. 하지만 기계 학습이 소프트웨어 개발의 주류에 접어들면서 비결정적 결과 도출이 더욱 보편화 될 것이다. 테스터로서 정답이 없는 애플리케이션 결과의 결함을 찾아 낸다고 하는 도전을 어떻게 받아들일 것인가?

### Bot보다 앞서 나아가야 한다.

앞으로 테스터가 AI Bot으로 대체 여부에 대해서 확신 하기는 어렵다. 하지만 한가지 분명한 것은 AI Bot의 도입으로 테스트 업무가 크게 변할 것이라는 것이다. 그 안에서 기계학습이 할 수 없는 사람만이 할 수 있는 역할을 찾아 적응의 길의 모색해야 할 것이다.

## 3 모바일 App 테스트 동향

본 내용은 Pipeline 2월호에서 발췌/번역한 기사로 작성자는 A1QA의 CIO, Dmitry Lubnevsky 입니다.

### 1. 기기의 단편화 증가 (Increasing Device Fragmentation)

오늘날 시장에는 다양한 유형 및 모바일 기기 모델이 출시되어 있으며 제조업체는 고객의 관심을 끌기 위해 끊임없이 새로운 기능을 개발하고 있다. 이것을 계속 유지하기 위해서 모바일 기기 테스트 공급 업체는 다음의 두 가지 방법을 사용할 수 있다. 하나는 물리적 모바일 기기 세트를 확장하는 것이고 다른 하나는 시장에서 사용 가능한 device farm을 사용하는 것이다. device farm은 자체적으로 물리적 장치를 사용하고 있어서 테스트하는 것과 마찬가지로 QA팀이 호스트 시스템을 통해 장치에 액세스 할 수 있다.

모바일 테스트 분야는 파레토 (Pareto) 법칙이라고 알려진 80/20 규칙을 적용하여 20 %의 기기에서 테스트를 실행하면 테스트 범위가 80 % 커버리지가 되는 효과를 가진다. 커버리지를 넓히려면 팀이 필요한 모든 기기 및 운영 체제 조합을 포괄 할 수 있도록 자동화 테스트가 필요할 수도 있다. 에뮬레이터로 테스트할 경우 비용이 덜 드는 것처럼 보일 수 있지만 에뮬레이터는 하드웨어를 고려하지 않으며 실시간 환경을 시뮬레이션 할 수 없다.

트렌드를 유지하고 수요가 많은 QA 공급 업체는 최대한 많은 기기에 대해 적절한 테스트 커버리지를 제공 가능하도록 준비해야 한다. 시장에서 사용되는 모든 기기에 대한 테스트를 수행하는 것은 불가능하며, 비용이 너무 많이 든다. 앱 테스트 계획을 수립 할 때 가능한 경우 핵심적인 통계치를 수집하고 최종 사용자 기기에서 직접 테스트를 실행하는 것이 중요하다. 또한 물리적으로 테스트 할 기기가 너무 많으면 device farm을 통한 자동화가 유일한 솔루션이다.

### 2. Chatbot의 성장

Chatbots은 응용 프로그램, 전자 메일 및 SMS에서 인간에게 응답, 질문 및 상호 작용을 시뮬레이트하는 대화형 소프트웨어이다. Google API, AI 및 IBM Watson 과 같은 프레임 워크를 사용하여 모바일 애플리케이션에서 Bot을 만드는 것이 매우 쉬워졌다. 오라클이 실시한 설문 조사에 따르면 2020년까지는 대화 상대의 80 %가 Chatbot이 될 것이라고 한다. 이것은 테스트 업계에 있어 새로운 기회가 될 것이다. 모든 chatbot의 목표는 유저의 삶을 편하게 하는 것이다. 그러나 Chatbot은 텍스트

또는 음성 입력에 의해 작동되며 이는 QA 및 개발 팀에게 새로운 과제가 된다.

테스터는 자연 언어 입력 상황을 인지할 수 있어야 하고, chatbot이 제공하는 정보가 적절한지 여부를 확인할 수 있어야 한다. 사용자 경험(UX) 시험은 chatbot의 확산에 주의를 요하는 또 다른 측면이다. 테스터는 chatbot이 사용자를 방해하거나 자극하지 않으며 관련 정보를 제공하기 위해 가능한 한 입력을 적게 요구하는지 확인해야 한다. 일부 조직에서는 chatbot을 앱 및 웹 사이트에 통합하는 것을 좋아하겠지만 테스터는 객관적이어야 하며 chatbot이 실제로 가치 제공을 하는지 확인해야 한다.

### 3. 음성 어시스턴트의 부상

현재 chatbot의 인기가 상승하고 있지만 기술은 이미 다른 커뮤니케이션 방법을 제공하기 위해 개발이 진행되고 있다. 텍스트 입력을 최소한으로 하고 음성으로 입력하는 시대가 올 것이다. Amazon의 Alexa, OK Google 및 Facebook Messenger는 유저와 소프트웨어 개발자의 상상력을 사로 잡고 있다. QA 테스터의 경우 음성 인식의 정확성을 보장하는 것이 중요하다. 이 기술은 모든 목소리, 사투리 및 악센트를 구별하여 사용자가 화면이 아닌 음성 인터페이스를 사용할 수 있도록 해야 한다. 스마트폰이나 웨어러블 기기의 화면이 작을수록 사용자는 음성에 더 많이 의존하게 되므로 모바일 기기에서 더욱 중요하게 사용될 것이며, 이 때 제대로 작동해야 한다. 유저 흐름을 점검할 수 있도록 철저한 테스트가 필요하다. QA 엔지니어는 화면을 터치 하지 않고 음성 어시스턴트를 호출 할 수 있는지 확인해야 한다. 그리고 인터럽트 상황을 고려한 테스트도 진행해야 한다. 테스터는 언어가 아닌 자연적인 소리, 예를 들면 기침을 하거나, 숨을 크게 들이쉬거나, 스마트폰 벨 소리 등에 대한 테스트도 진행해야 한다. 이런 것들을 앱에서 처리하기 쉽지 않으며 앱은 내가 무엇을 찾는 지 쉽게 이해 하기 어려운 문제이다. 이러한 모든 문제들을 전문 테스터들이 경험을 바탕으로 하여 해결해야 한다.

### 4. 모바일 결제

Android Pay, Apple Pay 및 Microsoft Wallet과 같은 혁신적인 모바일 결제 솔루션이 전 세계에 배포되어 유저는 스마트 기기를 결제 수단으로 사용할 수 있게 되었다. 2018년에는 PayPal과 같은 기존 온라인 지불 방법을 사용하지 않거나 사용하는 횟수가 줄어들게 될 것이다. 내부 독점 프로토콜을 기반으로 하는 Samsung Pay, Microsoft Wallet, Apple pay 등과 같은 새로운 유형의 지불 옵션이 더 자주 사용

될 것이다. 결제 게이트웨이 테스트에는 기능 테스트, 통합 테스트 및 심오한 보안 테스트가 포함되어 있어야 한다.

## 5. 보안 강조

오늘날 사람들은 스마트 폰을 어느 때 보다 자주 사용한다. 우리는 스마트폰을 이용하여 인터넷 서핑을 하거나 친구들과 채팅을 한다. 앱의 가능성이 빠른 속도로 발전하고 있으며 **최종 유저와 개인 데이터에 대해 충분한 수준의 보안을 제공해야만** 한다. 오늘날 사람들은 점점 더 보안 문제에 대해 인식하고 있어서 의심스러운 링크를 클릭하거나 확인되지 않은 데이터를 다운로드 하지 않는다. 그러나 테스터는 사용자 관행에만 의존 할 수 없다. 최근 OWASP는 **보안 코딩 연습 가이드**라는 개발 라이프 사이클에 적용 할 수 있는 포괄적 인 체크리스트를 발행했다. 이것은 **모바일 테스트 도메인을 위한 고유 한 보안 정책의 필수 단계**이다. 보안 테스트 지식과 기술은 모바일 테스트와는 차별화 된 요소이다.

## 6. 가상 현실과 증강 현실

VR과 AR은 소리, 감각, 청력 및 촉각을 시뮬레이팅하는 인공 환경을 만드는 기술이다. 이 기술은 매우 빠르게 성장하고 있으며 이 성장으로 테스트를 위한 틈새 시장이 늘어나고 있다.

Google, Microsoft 또는 Facebook과 같은 기업은 VR / AR에 막대한 투자를 하고 있으며 이러한 기술을 가지고 있는 신생 기업을 인수하고 있다. QA 업체는 모바일 게임 산업이 VR / AR 기술의 가장 큰 영역이므로 모바일 테스트 경험을 기반으로 VR 테스트를 하려는 경향이 있지만 VR / AR에는 특유의 고려사항이 있다.

모든 VR / AR 응용 프로그램은 고유 한 논리를 가지고 있으며 매우 다양하다. 그러나 기술의 본질이 비기능적 테스트를 기능적 테스트만큼 중요하게 만든다는 것은 확실하다. UX 테스트 전문가는 **소프트웨어 사용을 추적하고 유저의 편의에 주의를 기울여야 하며 유저의 피로를 최소화해야 한다**. 테스트 중인 VR 앱을 사용하는 동안 유저가 어지럽거나 아프지 않을 거라고 확신 할 수 있는가? 이를 보장하기 위해서 **경험을 통한 광범위한 테스트가 필요하다**. 최고의 경험을 유저가 보장받기 위해서 모든 테스터는 유저가 편안하게 앱을 사용할 수 있도록 장치 별 교육을 받아야 한다.

VR 앱은 대부분의 고객에게 새로운 기능이므로 유저 오류가 전체 앱의 사용 및 다른 유저의 경험에 영향을 줄 수 있는지도 확인해야 한다.



## 7. 접근성

앱은 사람들이 정부 및 주 교육 기관과 상호 작용할 수 있는 수단이 되어가고 있다. 정부가 개발하거나 지원하는 모바일 앱은 이미 주차장 찾기 및 요금 지불, 시민들의 자동차 공유 또는 폐기물 재활용 참여와 같은 다양한 서비스 제공 기회를 제공하고 있다. 앱을 통해 주민에게 다양한 알릴 사항, 긴급 알림 등을 할 수 있다. 다양한 성격에도 불구하고 이러한 모든 앱에는 공통점이 하나 있다. 장애가 있는 유저가 쉽게 접근 가능해야 한다는 것이다.

모든 사용자에게 강력하고 다양한 응용 프로그램을 제공하기 위해 고려해야 할 표준 및 규정이 있다. 소프트웨어 테스터는 모두가 사용할 수 있도록 책임을 가지고 테스트에 임해야 한다.

## 8. 필수적인 도메인 지식 습득

오늘날 모바일 앱은 BFSI, 건강 관리, 게임, 전자 상거래 등 모든 도메인을 지원하고 있다. 테스터는 각 도메인의 특성을 완벽하게 이해해야 한다. 기능 및 보안 문제에 대해서만 앱을 테스트하는 것만으로 충분하지 않다. 테스터는 고유한 비즈니스, 업계 및 유저의 요구 사항을 파악하고 그에 따라 앱을 테스트해야 한다. 팀에 프로젝트를 할당하기 전에 팀이 비즈니스 도메인을 파악하고 앱이 고유한 환경에서 어떻게 움직여야 하는지 확인해야 한다.

### 결론

디지털 기술과 모바일 기술의 급격한 변화에 따라 기업은 QA 업체에게 전문적인 모바일 앱 테스트 접근 방법을 기대하고 있다. 소프트웨어 테스터는 다양한 기기 및 운영체제를 이해하고 있어야 하며 앱이 개발된 용도에 맞는 지 확인해야 한다.

사회 개념의 측면과 모바일 앱의 가능성에 대해 주목하고 있다는 점도 소프트웨어 구축 방식을 변화시키고 있다. 앞으로의 추세에 주목하여 QA 및 테스트 조직은 포괄적인 테스트 접근 방식을 채택하고 비즈니스 및 고객을 위해 강력한 모바일 제품을 제공할 수 있어야 한다.