

L1C6 - Use Case: Global Health

0:01 in the last two concepts we walked
0:03 through two different use cases the
0:04 first was a global trade use case and
0:06 the second was a money laundry use case
0:09 and in both of these use cases we showed
0:11 how existing data science tools and
0:13 techniques failed to bring together
0:15 enough data to solve these really really
0:17 important challenges but that through
0:20 privacy ends in techniques and in
0:21 federated data networks a solution could
0:24 be met and in the third use case what
0:26 we're really going to see is so the same
0:28 core idea we're going to shift over to
0:30 to global health as as and this is one
0:32 of the most obvious motivating use cases
0:34 of privacy handsy technologies just
0:36 because in context of global health we
0:38 know that we want to study patterns
0:40 across lots of different humans these
0:41 these often are very very personal
0:43 obviously to each individual but also
0:45 really really important to to be able to
0:47 solve for the benefit of society
0:49 um and the the there's many use cases in
0:52 the global health spectrum that we could
0:53 use as examples um but the one we want
0:55 to focus on here is is is breast cancer
0:58 and the reason we're gonna focus on
0:59 breast cancer is it's another example of
1:02 of something that that really
1:03 really should be open and shut in many
1:05 ways specifically identifying breast
1:07 cancer but it isn't because the data is
1:09 so
1:10 so separated so so divided amongst so
1:13 many different organizations around the
1:14 world and very few people are able to
1:16 really aggregate meaningful amounts of
1:18 it and thus far no one has been able to
1:20 aggregate enough to really solve this
1:22 important challenge so
1:23 first a little bit of background
1:24 information on on breast cancer
1:27 specifically
1:29 breast cancer is treatable when caught
1:31 early right but hundreds of thousands of
1:32 people still die from it every year so
1:34 so an estimated 2.2 million incidents of
1:37 breast cancer
1:38 in 2020 and and just under 700 000
1:42 people passed away um and globally
1:44 there's this really well-known statistic
1:45 that roughly one in eight women suffer
1:47 from breast cancer at some point in in
1:49 their life
1:50 and the the challenging aspect is
1:52 despite the fact that this is such a
1:54 a widespread
1:57 disease

私たちが歩んできた最後の 2 つの概念において
2 つの異なるユースケースを通じて、
最初は**世界貿易**のユースケースで、
2 つ目は**マネー ランドリー**の使用例でした
これらの両方のユースケースで私たちが示したのは、
既存のデータサイエンスツールと
テクニックを組み合わせることができなかった
これらを本当に本当に解決するのに十分なデータ
重要な課題ですが、
プライバシーは技術で終わる
ソリューションが実現できるフェデレーション データ ネットワーク
3番目のユースケースでは何が満たされますか
私たちは本当に同じであることがわかります
これから移行する**中心となるアイデア**
世界の健康に向けて、これもその 1 つです
最も明らかな動機付けとなるユースケース
プライバシーの便利な**テクノロジー**だけで
なぜなら、世界的な健康の文脈において、私たちは
パターンを研究したいのはわかっています
多くの異なる人間にわたって、これらは
これらは多くの場合非常に個人的なものです
もちろん人それぞれですが、
できるようになることが**本当に本当に重要**
社会の利益のために解決する
ええと、使用例はたくさんあります
私たちができる世界的な健康スペクトル
例として使用しますが、私たちが望んでいるのは
ここで焦点を当てるのは乳がんです
そして私たちが注目する理由
乳がんもその一例です
本当に何かのこと
本当に多くのところで開閉する必要があります
乳房を具体的に特定する方法
癌ではありますが、データがあるため癌ではありません
それで
とても離れていて、とても分裂しています
周囲のさまざまな組織
世界中でそれができる人はごくわずかです
本当に意味のある量の
そして今のところ誰もそれができていない
これを実際に**解決するのに十分な量を集約する**
重要な課題なので
最初に背景を少し説明します
乳がんに関する情報
具体的には
乳がんは発見されれば治療可能
早い段階で右ですが、数十万
今でも毎年それで人が亡くなっているので、
したがって、推定220万件のインシデントが発生します。
乳癌
2020年には70万人弱
世界中で人々が亡くなりました
これは非常によく知られた統計です
およそ8人に1人の女性が苦しんでいる
ある時点で乳がんになった
彼らの人生
そして挑戦的な側面は
これがそのようなものであるにもかかわらず、
広く普及した
疾患

1:58 that that we still have a struggle
2:01 actually identify right in and that that
2:04 mammographies or mammograms have a very
2:06 very significant accuracy problem in
2:08 that that if you go in for for for a
2:10 scan to see if you have a breast cancer
2:13 there's a roughly one in four false
2:15 positive and false negative rate of that
2:18 experience meaning that the one in four
2:20 people who actually have breast cancer
2:23 leave the clinic being told that that
2:25 they don't right
2:27 and ai machine learning the the the
2:29 latest models put into production are
2:31 even worse than the typical radiologist
2:33 right closer to
2:34 to one and three so divide the fact that
2:36 that that this is a global
2:39 um important
2:41 deadly disease that is well understood
2:43 there's lots of money behind lots of
2:44 infrastructure lots of people getting
2:46 tested the the act of getting tested has
2:48 a serious serious accuracy problem and
2:50 what's even more challenging is it it's
2:51 such a global response like so many
2:54 people are getting tested so frequently
2:56 that there's there's an enormous amount
2:58 of data that exists you know we estimate
3:00 that around 800
3:02 million mammography images are taken
3:04 every single year
3:06 but
3:07 in terms of the size of data sets
3:09 actually used to train an ai or machine
3:11 learning model you know it's less than
3:14 0.1 percent of the world's mammography
3:17 images produced over over the last 10
3:19 years and so when we think about you
3:20 know the the what **imagenet** for example
3:23 did to to
3:24 image classification accuracy like like
3:26 several orders of magnitude increase in
3:28 the amount of data that was available
3:30 led to a huge increase and ultimately
3:33 inaccuracy and ultimately surpassing the
3:35 accuracy of a human annotator and this
3:38 is what we want to see for for
3:41 for being able to model breast cancer
3:42 and be able to detect breast cancer
3:44 because again if it's if you catch it
3:46 early if we can we can lower the cost
3:47 and increase the accuracy of detecting
3:49 breast cancer we can save thousands and
3:51 thousands and thousands of lives every
3:52 year
3:53 and
3:54 and
3:55 the **the challenge is that no one has**
3:56 enough data **right in the sense** that that
3:58 you know if there's if there's nearly
4:00 eight billion images in existence and
4:02 and the production models that are going
4:04 out are only trained at a million images

私たちはまだ闘いがあるということ
実際にそれを特定します
マンモグラフィーやマンモグラム検査には非常に大きな特徴があります。
非常に重大な精度の問題
あなたがのために行ったら、それは
乳がんがあるかどうかを調べるスキャン
およそ 4 人に 1 人が**虚偽**です
その**陽性率と偽陰性率**
経験は4人に1人を意味します
実際に乳がんを患っている人
そう言われてクリニックを後にする
彼らは正しくない
そしてAI機械学習は、
生産されている最新モデルは、
典型的な放射線科医よりもさらに**悪い**
すぐ近くにある
という事実を 1 と 3 に分割します。
これがグローバルだということ
えー、重要です
よく理解されている致命的な病気
たくさんのお金がある
多くの人がインフラストラクチャを取得
テストされた、テストを受けるという行為は
重大な精度の問題と
さらに難しいのは、それです
多くの人々と同様に、このような**世界的な**反応
人々は頻繁に検査を受けています
膨大な量があるということ
ご存知のとおり、存在するデータのうちの 1 つを私たちが推定しています
それは約800
マンモグラフィー画像は100万枚撮影される
毎年
しかし
データセットのサイズの観点から
AIや機械を訓練するために**実際に使用される**
学習モデルはそれより小さいことがわかっています
世界のマンモグラフィーの 0.1%
過去 10 年間に作成された画像
何年も、**私たちがあなたのことを考えるとき**
たとえば、**imagenet** が何であるかを知っています
したこと
画像分類精度 同様
数桁の増加
利用可能なデータの量
それが大幅な増加につながり、最終的には
不正確であり、最終的には基準を超えています。
ヒューマン・アノテーターの精度とこれ
私たちが見たいのは
乳がんのモデル化ができたことに対して
そして乳がんを発見できるようになります
なぜなら、もしそれがキャッチできたらまたそうなるから
できれば早めにコストを下げることができます
検出の精度を高めます
乳がんは何千人も救えるし、
毎年何千もの命が
年
と
と
課題は誰も持っていないことだ
その意味で十分なデータがある
ほぼあるかどうかはわかります
80 億の画像が存在し、
そして今後の量産モデル
out は 100 万枚の画像でのみトレーニングされます

4:05 then then
4:07 this is
4:08 you know it's a partnership problem not
4:10 a machine learning process not
4:11 necessarily a modeling problem it's not
4:13 that our classifiers need to get that
4:14 much better it's just that there's a
4:16 huge latent resource that we don't have
4:18 the ability to have access to you know
4:19 so when we ask this question why can't
4:21 ai perfectly predict breast cancer you
4:23 know it's because the world's hospitals
4:25 can't or don't put all their medical
4:27 data into one place to train a cancer
4:30 classifier right in a sense that there
4:32 are thousands and thousands and
4:34 thousands literally thousands of
4:36 mammography screening clinics around the
4:37 world um but because they're all siloed
4:41 in their own individual data silos
4:43 identifying breast cancer is still an
4:45 incredibly low precision and low recall
4:47 activity
4:48 whether it's an ai or or a an individual
4:51 participating in it and
4:53 this brings us back to our **you know our**
4:55 **pets theme right privacy answer**
4:56 **technologies make it possible to answer**
4:58 questions using data that you cannot see
5:00 using data that you don't have access to
5:02 that can be distributed across thousands
5:04 of thousands of different organizations
5:06 and so when we encounter a question like
5:08 why can't ai perfectly predict breast
5:10 cancer when there is such an enormous
5:13 amount of data that exists on this
5:15 problem and there's such an enormous
5:17 amount of need one in eight women in the
5:19 world get breast cancer at some point in
5:21 their life we have a huge resource we
5:23 have a huge need but the problem doesn't
5:25 get solved and the answer
5:27 is because the data is distributed
5:28 across thousands and thousands of
5:30 different organizations
5:31 and this is the promise of privacy
5:33 enhanced technology it is you know
5:35 another example of something that is you
5:37 know
5:38 it is open and shut that this should be
5:41 solvable by machine learning we have an
5:43 enormous amount of data an enormous
5:44 amount of need an enormous amount of
5:46 funding that goes towards towards
5:48 tackling breast cancer
5:49 we know that we can detect it early
5:52 sorry we know that if we detected early
5:53 that it would save a lot of lives and we
5:55 have really powerful machine learning
5:57 and ai tools that should be able to
5:58 detect it early but they just don't have
6:00 access to the huge latent resources
6:02 necessary to make it happen
6:05 and so i hope across these three

それから
これは
それはパートナーシップの問題ではないことはわかっています
機械学習プロセスではありません
必ずしもモデリングの問題ではありません
私たちの分類器はそれ取得する必要がある
があるだけでずっと良いのです。
私たちが持っていない巨大な潜在的リソース
あなたが知っているものにアクセスする能力
では、この質問をすると、なぜできないのでしょうか？
AI が**あなたの乳がん**を完璧に予測します
それは**世界の病院のせいだ**ということを知ってください
医療費をすべて投入できない、または投入できない
データを 1 か所に集めてがんを訓練する
ある意味での分類子
何千も何千もの
何千も文字通り何千もの
周辺のマンモグラフィ検査クリニック
世界、うーん、でもそれらはすべてサイロ化されているので
独自のデータサイロ内にある
乳がんの特定は依然として難しい
信じられないほど低い精度と低い再現率
アクティビティ
AIであろうと個人であろうと
それに参加して、
これは私たちを私たちに戻します、あなたが知っている私たちのこと
ペットのテーマの正しいプライバシーの答え
テクノロジーが答えを可能にする
目に見えないデータを使った質問
アクセス権のないデータを使用する
何千人にも分散できる
何千もの異なる組織の
そして、次のような質問に遭遇したとき、
なぜAIは胸を完全に予測できないのか
こんなに巨大な癌があると
この上に存在するデータの量
問題は非常に大きな問題です
女性の8人に1人が必要とする量
世界のある時点で乳がんになる
彼らの人生、私たちには**膨大な資源がある**
大きなニーズがある**が、問題は解決していない**
解決して答えを得る
データが分散されているからです
何千もの
さまざまな組織
そしてこれはプライバシーの約束です
強化されたテクノロジーです
あなたというものの**別の例**
知る
これは開いた状態と閉じた状態であるべきです
機械学習によって解決できる
膨大な量のデータ
必要な量 膨大な量
に向けて使われる資金
乳がんとの闘い
私たちはそれを早期に発見できることを知っています
申し訳ありませんが、早期に発見できていればわかります
それが多くの命を救うだろうし、私たちも
非常に強力な機械学習を備えている
そしてできるはずのAIツール
早期に発見するが、彼らはそれを持っていない
膨大な潜在リソースへのアクセス
それを実現するために必要な
だから**私はこの3つを乗り越えることを願っています**

6:06 different use cases you see this same
6:08 template you see the same theme which is
6:10 the world is big it has
6:13 so many organizations it has so many
6:15 people and even inside those
6:16 organizations it has so many different
6:18 departments you know that are that even
6:20 themselves can be distributed around the
6:21 world and and
6:23 okay yes the internet is big and we can
6:25 scrape data and this kind of stuff but
6:27 like
6:28 the
6:29 the really valuable data the stuff about
6:31 what's going on in the world the stuff
6:32 about what's going on in our lives this
6:35 is is not widely available but there is
6:37 a huge amount of it right and and even
6:40 if you know in the case of breast cancer
6:42 you could you could
6:44 you could form a partnership with a
6:45 hospital maybe two hospitals maybe maybe
6:48 three or five or ten hospitals but
6:51 this data is spread across thousands of
6:54 organizations
6:56 it's just it'd be it's totally
6:57 unprecedented i don't want to say that
6:59 it's impossible but the the amount of
7:00 cost that it would require for you as a
7:02 data scientist to enter into thousands
7:05 of individual data sharing agreements
7:07 all the partnership and risk and legal
7:09 and the finances how much money you'd
7:10 have to pay to acquire a copy of all
7:13 this personal data right
7:15 it's just such an insurmountable
7:18 challenge such a huge mountain to to to
7:21 climb
7:22 and privacy-enhanced technologies are
7:24 going to create federated data networks
7:26 that make this
7:28 just so much easier and you know i'm not
7:30 going to get into exactly how yet but
7:32 but this is the promise that we're
7:33 talking about in in this field um and
7:36 and you know we could we could continue
7:38 with more use cases like there's there's
7:39 many more you know we talked about trade
7:40 we talked about money laundering we
7:41 could talk about you know national
7:43 security or or intelligence or
7:44 disinformation or fairness like lots and
7:46 lots of important problems where where
7:48 the data to solve them exists
7:52 but the implications of centralizing it
7:54 or the cost of centralizing or the cost
7:56 of acquiring is simply too high friction
7:59 for anyone to have legitimately solved
8:01 that problem right and you know i i'm
8:03 including a big list here some of these
8:04 **include you know various governance**
8:06 **problems** of you know who's controlling
8:08 when people ought to answer these
8:09 questions you know we're we're we're

異なるユースケースでも同じことがわかります
テンプレートと同じテーマが表示されます。
世界は広いです
非常に多くの組織があります
人々、そしてその内部でも
非常に多くの異なる組織があります
あなたが知っている部門はそれさえあります
それ自体を周囲に分散させることができます。
世界として
はい、インターネットは大きいので、それが可能です
データをスクレイピングしたりすることもあります、
好き
の
本当に貴重なデータに関するもの
世界で何が起こっているのか
私たちの人生で何が起こっているのかについて
あまり広くは入手できませんが、
大量のそれが正しく、そして均等に
乳がんの場合に知っていれば
できるかもしれない、できるかもしれない
〜とパートナーシップを築くことができます
病院かもしれない 2 つの病院かもしれない かもしれない
3つ、5つ、10の病院だけど、
このデータは何千ものデータに分散されています
組織
それはただそれだろう それは完全に
前例のないことは言いたくない
無理だけどその量は
として必要となる費用
データサイエンティストが数千人に入る
個々のデータ共有契約の
すべてのパートナーシップ、リスク、法的責任
そして財政はいくらですか
すべてのコピーを入手するには料金を支払わなければなりません
この個人データの権利
それはまさに乗り越えられないことだ
あんなに大きな山に挑戦して〜
登る
プライバシー強化テクノロジーは、
フェデレーション データ ネットワークを作成する予定です
これを作るのは
とても簡単です、そしてあなたは私がそうではないことを知っています
正確にはまだ説明するつもりですが、
でもこれが私たちの約束です
この分野について話しているのですが、ええと、
そして、あなたは知っています、私たちは続けることができるでしょう
他にもユースケースはたくさんあります
私たちが貿易について話したのはご存知のとおりです
私たちはマネーロンダリングについて話しました
あなたが全国的に知っていることについて話すことができます
セキュリティまたはインテリジェンスまたは
多くのような偽情報や公平性
多くの重要な問題がどこにあるのか
それらを解決するデータは存在します
しかし、それを一元化することの影響
または集中化のコストまたはコスト
取得の摩擦が高すぎる
誰でも合法的に解決できるように
その問題は正しい、そしてあなたは私がそうであることを知っています
ここに大きなリストが含まれています。
さまざまな**ガバナンスを知っている**ことも含めて
誰がコントロールしているか知っているという**問題**
人々がこれらに答えるべきとき
質問、あなたは知っていますか？

8:10 stepping into some of **society's biggest**
8:13 **biggest issues** but but the tools that
8:15 we're stepping in with to **solve these**
8:17 **issues are so much more powerful** than
8:20 the ones we're leaving behind that that
8:22 um
8:23 that i really hope you can see the
8:25 promise of what this sort of next
8:27 generation of the internet this this
8:28 next generation of ais next generation
8:30 of data science this next generation of
8:32 statistics and machine learning is is
8:34 going to bring
8:35 um so um this is this has been the the
8:38 **third use case and last one recovered**
8:39 from the next concept we're really going
8:41 to wrap up these three different use
8:42 cases with what they have **in common** and
8:44 transition into talking about the the
8:46 underlying technologies that are going
8:47 to be able to solve them
8:49 see in the next concept
英語 (自動生成)

社会最大の組織に足を踏み入れる
最大の問題はツールですが、
私たちはこれらを解決するために協力しています
問題はそれよりもはるかに強力です
私たちが残していくもの
えーっと
あなたが見てくれることを本当に願っています
次はこういうことを約束する
インターネットの世代 これこれ
次世代AIS次世代
次世代のデータサイエンスの
統計と機械学習は
持っていくつもりです
ええと、ええと、これはこれでした
3 番目のユースケースと最後のユースケースが回復されました
次のコンセプトから本当に進んでいきます
これら 3 つの異なる使用法をまとめると、
共通点があるケースと、
についての話に移ります
今後の基盤となるテクノロジー
それらを解決できるように
次のコンセプトを参照してください



OpenMined

<https://www.youtube.com/watch?v=iAr98BYtsyQ>